

Avaliação de QoS/QoE no Serviço de Videochamada Skype™

André Duarte Lopes

Uma dissertação submetida à Universidade do Minho para obtenção do grau de Mestre em Engenharia de Comunicações elaborada sob a orientação do Professor Paulo Martins de Carvalho.

Universidade do Minho

Escola de Engenharia

Outubro, 2012

Agradecimentos

Gostaria de agradecer ao Professor Paulo Martins de Carvalho pela constante orientação e disponibilização de recursos. Gostaria de agradecer também à Professora Solange Rito Lima pelo seu auxílio e contribuição de ideias.

Finalmente um agradecimento à minha família e a todos que contribuíram com o seu apoio.

Abstract

Web-based videotelephony services currently occupy a prominent place in the wide range of services and generated network traffic. With the growing use of such services comes also an increasing need for the evaluation of user experience. Depending on network conditions, there are relevant QoS parameters (bandwidth, delay, loss ratio, jitter, etc.) that have an impact on the quality of experience of the service.

The objective of this thesis is to find out which network parameters affect videocalls quality of experience, which are the threshold values for which the quality of experience levels are affected, in what way is the quality of experience impacted and what is their relation.

The quality of experience was evaluated using objective methods for all video and audio samples collected during the experimentation phase. Conclusions were made based on the results which are presented in this dissertation.

Resumo

Os serviços de vídeo-telefonia na rede Internet ocupam actualmente um lugar de relevo no leque de serviços e tráfego de rede gerado. Com a crescente utilização deste tipo de serviços surge também uma maior necessidade de avaliar a qualidade de experiência que o utilizador experiencia. Dependendo do tipo de condições que uma rede apresenta, existem parâmetros de QoS relevantes (largura de banda, atraso, perda de pacotes, jitter, etc.) que afectam a qualidade de experiência do uso do serviço.

Esta dissertação tem como objectivo descobrir quais os parâmetros de uma rede têm efeito na qualidade de experiência de uma videochamada, quais as condições limite para as quais os níveis de qualidade de experiência são afectados, de que forma é a qualidade de experiência afectada e qual a sua relação.

A qualidade de experiência foi avaliada usando métodos objectivos em todas as sequências vídeo e áudio recolhidas durante a fase de testes. Foram retiradas conclusões com base nos resultados obtidos que são apresentadas nesta dissertação.

Conteúdo

Agradecimentos	iii
Abstract	v
Resumo	vii
Conteúdo	ix
Lista de Figuras	xiii
Lista de Tabelas	xvii
Lista de Acrónimos	xix
1 Introdução	1
1.1 Introdução	1
1.2 Motivação e Objetivos	2
1.3 Metodologia de Estudo	3
1.4 Organização da Dissertação	3
2 Estado da Arte	5
2.1 De Qualidade de Serviço a Qualidade de Experiência	5
2.2 Qualidade de Experiência	7

CONTEÚDO

2.3	Mean Opinion Score	11
2.4	Peak Signal-to-Noise Ratio	12
2.5	Perceptual Evaluation of Speech Quality	13
2.6	NTIA General Model - VQM	15
2.6.1	Técnicas de Calibração	16
2.6.2	Características e Parâmetros do General Model	19
2.6.3	Cálculo e Desempenho do General Model	20
2.7	Resumo	24
3	Métodos para Análise de Vídeo	25
3.1	Batch Video Quality Metric	25
3.2	Método <i>Full Reference</i> para estimar atrasos de vídeo	29
3.3	VQM_VFD	31
3.3.1	Parâmetros VFD	31
3.3.1.1	Parâmetro 1 (Par1)	32
3.3.1.2	Parâmetro 2 (Par2)	34
3.3.1.3	Correlação dos parâmetros VFD	35
3.3.2	Video Quality Model para Variable Frame Delay (VQM_VFD)	37
3.3.2.1	Parâmetros VFD baseados na percepção	38
3.3.2.2	Correlação do VQM_VFD	41
3.4	Resumo	44
4	Arquitetura e Metodologia de testes	45
4.1	Arquitetura do Sistema	45
4.2	Emissão e Recepção de Vídeo/Áudio	46
4.3	Sequências de Vídeo e Áudio usadas	47
4.4	Análise dos dados	49

4.5	Resumo	49
5	Avaliação de QoE/QoS no Serviço de Videochamada	51
5.1	Loss Ratio vs Delay	53
5.2	Loss Ratio vs Bandwidth	55
5.3	Bandwidth vs Delay	58
5.4	Jitter	60
5.5	TCP vs UDP	63
5.6	Resumo	66
6	Conclusões	67
6.1	Trabalho Futuro	69
	Bibliografia	71
	Anexo	77
A	Resultados individuais da avaliação de vídeo	77
A.1	Loss Ratio vs Delay	77
A.2	Loss Ratio vs Bandwidth	78
A.3	Bandwidth vs Delay	78
A.4	Jitter	79
A.5	TCP	80
A.6	UDP	81
B	Resultados individuais da avaliação de áudio	83
B.1	Loss Ratio vs Delay	83
B.2	Loss Ratio vs Bandwidth	84
B.3	Bandwidth vs Delay	84

CONTEÚDO

B.4	Jitter	85
B.5	TCP	86
B.6	UDP	87

Lista de Figuras

2.1	QoE/QoS management life-cycle [1]	6
2.2	Forma geral da curva que mapeia a relação QoS/QoE [2]	7
2.3	Influências na experiência do utilizador - Forlizzi & Ford [3]	8
2.4	Avaliação da qualidade geral da chamada	9
2.5	Identificação de problemas ocorridos	10
2.6	Esquema geral do funcionamento do PESQ ([4] Fig. 1)	14
2.7	Diagrama de blocos do VQM ([5] Fig.1)	16
2.8	Qualidade subjectiva de vídeo vs VQM ([5] Fig. 2)	22
2.9	Qualidade subjectiva de vídeo HRC vs HRC VQM ([5] Fig. 3)	22
2.10	VQEG FR-TV Phase II - 525 linhas: Qualidade subjectiva de vídeo vs VQM ([5] Fig. 4)	23
2.11	VQEG FR-TV Phase II - 625 linhas: Qualidade subjectiva de vídeo vs VQM ([5] Fig. 5)	23
3.1	Interface gráfico BVQM	26
3.2	Qualidade subjectiva vs traçado objectivo para Par1 e Par2 ([6] Fig. 2)	36
3.3	Esquema geral do VQM_VFD ([7] Fig. 1)	37
3.4	Píxeis contidos num bloco ST ([7] Fig. 2)	38
3.5	Desempenho do modelo VQM_VFD ([7] Fig. 5)	43
4.1	Arquitectura do sistema de testes	46

LISTA DE FIGURAS

4.2	Captura de ecrã do vídeo "chart"	48
4.3	Captura de ecrã do vídeo "train"	48
5.1	Qualidade vídeo "chart" - Loss Ratio vs Delay	54
5.2	Qualidade áudio "chart" - Loss Ratio vs Delay	54
5.3	Qualidade vídeo "train" - Loss Ratio vs Delay	55
5.4	Qualidade áudio "train" - Loss Ratio vs Delay	55
5.5	Qualidade vídeo "chart" - Loss Ratio vs Bandwidth	56
5.6	Qualidade áudio "chart" - Loss Ratio vs Bandwidth	56
5.7	Qualidade vídeo "train" - Loss Ratio vs Bandwidth	57
5.8	Qualidade áudio "train" - Loss Ratio vs Bandwidth	57
5.9	Qualidade vídeo "chart" - Bandwidth vs Delay	59
5.10	Qualidade áudio "chart" - Bandwidth vs Delay	59
5.11	Qualidade vídeo "train" - Bandwidth vs Delay	60
5.12	Qualidade áudio "train" - Bandwidth vs Delay	60
5.13	Qualidade vídeo "chart" - Jitter	62
5.14	Qualidade áudio "chart" - Jitter	62
5.15	Qualidade vídeo "train" - Jitter	63
5.16	Qualidade áudio "train" - Jitter	63
5.17	Qualidade vídeo "chart" - TCP vs UDP	65
5.18	Qualidade áudio "chart" - TCP vs UDP	65
5.19	Qualidade vídeo "train" - TCP vs UDP	66
5.20	Qualidade áudio "train" - TCP vs UDP	66
A.1	Vídeos "chart" - Pontuações VQM_VFD - LR vs Delay	77
A.2	Vídeos "train" - Pontuações VQM_VFD - LR vs Delay	77
A.3	Vídeos "chart" - Pontuações VQM_VFD - LR vs BW	78

A.4	Vídeos "train" - Pontuações VQM_VFD - LR vs BW	78
A.5	Vídeos "chart" - Pontuações VQM_VFD - BW vs Delay	78
A.6	Vídeos "train" - Pontuações VQM_VFD - BW vs Delay	79
A.7	Vídeos "chart" - Pontuações VQM_VFD e medições de jitter	79
A.8	Vídeos "train" - Pontuações VQM_VFD e medições de jitter	79
A.9	Vídeos "chart" - Pontuações VQM_VFD - TCP	80
A.10	Vídeos "train" - Pontuações VQM_VFD - TCP	80
A.11	Vídeos "chart" - Pontuações VQM_VFD - UDP	81
A.12	Vídeos "train" - Pontuações VQM_VFD - UDP	81
B.1	Faixas áudio "chart" - Pontuações MOS - LR vs Delay	83
B.2	Faixas áudio "train" - Pontuações MOS - LR vs Delay	83
B.3	Faixas áudio "chart" - Pontuações MOS - LR vs BW	84
B.4	Faixas áudio "train" - Pontuações MOS - LR vs BW	84
B.5	Faixas áudio "chart" - Pontuações MOS - BW vs Delay	84
B.6	Faixas áudio "train" - Pontuações MOS - BW vs Delay	85
B.7	Faixas áudio "chart" - Pontuações MOS - Jitter	85
B.8	Faixas áudio "train" - Pontuações MOS - Jitter	85
B.9	Faixas áudio "chart" - Pontuações MOS - TCP	86
B.10	Faixas áudio "train" - Pontuações MOS - TCP	86
B.11	Faixas áudio "chart" - Pontuações MOS - UDP	87
B.12	Faixas áudio "train" - Pontuações MOS - UDP	87

LISTA DE FIGURAS

Lista de Tabelas

2.1	Mean Opinion Score	11
5.1	Tipologia dos testes	52

LISTA DE TABELAS

Lista de Acrónimos

AVI	Audio Video Interleave
BVQM	Batch Video Quality Metric
DMOS	Difference Mean Opinion Score
DSCQS	Double Stimulus Continuous Quality Scale
FR	Full Reference
IPTV	Internet Protocol Television
ITU-T	International Telecommunication Union Telecommunication Standardization Sector
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MSE	Mean Squared Error
NTIA	National Telecommunications and Information Administration
PCM	Pulse-code Modulation
PESQ	Perceptual Evaluation of Speech Quality
PEVQ	Perceptual Evaluation of Video Quality
PSNR	Peak Signal-to-Noise Ratio
ST	Spatial-Temporal
QoE	Quality of Experience
QoS	Quality of Service
SLS	Service Level Specification

LISTA DE ACRÓNIMOS

TCP	Transmission Control Protocol
UDP	User Datagram Protocol
VFD	Variable Frame Delay
VQM	Video Quality Model

Capítulo 1

Introdução

1.1 Introdução

Os serviços de telefonia ou vídeo-telefonia na rede Internet ocupam actualmente um lugar de relevo no leque de serviços e tráfego de rede gerado. Em 2009 a aplicação SkypeTM foi responsável por 12% do total de minutos de chamadas internacionais, em 2010 esta percentagem cresceu para 25% do total de minutos em chamadas internacionais sendo neste momento, destacadamente, o líder em comunicações internacionais de voz [8]. As videochamadas representam já uma fatia considerável da utilização destas aplicações, em minutos gastos em chamadas Skype-to-Skype em 2010, 42% foram videochamadas [9].

Existe uma grande variedade de aplicações a oferecerem serviços de VoIP/videochamada. Com este tipo de aplicações surge também uma maior atenção ao estudo de como avaliar a qualidade do conteúdo que é entregue ao utilizador. Dependendo do tipo de aplicações/serviços e dos parâmetros de Qualidade de Serviço (QoS) considerados relevantes (largura de banda, atraso fim-a-fim, perda de pacotes, etc.), existem limites qualitativos e/ou quantitativos para os quais os utilizadores percebem uma Qualidade de Experiência (QoE) variada do uso do serviço.

Os parâmetros de QoS e respectivos limites, relevantes para os fornecedores de serviço na especificação de serviços, configuração e monitorização dos recursos de rede, são de limitado interesse para o utilizador final. Daí que se torna relevante o estudo de métricas de QoE associadas a este serviço.

A qualidade de experiência determina o valor percebido resultante da utilização de um serviço ou aplicação por um utilizador e pode ser determinante para a escolha de uma aplicação

em detrimento de outra. A necessidade da indústria para obtenção de métricas de avaliação de vídeo tem aumentado com aplicações e serviços como IPTV, vídeo na Internet e crescente utilização de smartphones com ligação à Internet [10].

O valor numérico mais utilizado e aceite para avaliar a qualidade de experiência de um serviço é o Mean Opinion Score (MOS), uma pontuação resultante de um método que consiste em apresentar um questionário a um número suficiente de utilizadores para que estes possam avaliar numericamente a experiência que tiveram. Uma avaliação MOS é no entanto bastante difícil de se efectuar, sendo um processo lento, e simplesmente impraticável em várias situações. Uma aproximação ao MOS pode ser no entanto obtida através de métodos objectivos e de medições concretas e fáceis de obter (PESQ [4], PEVQ [11]) podendo estas medições serem traduzidas em diferentes modelos representativos de QoE (tais como [12][13][14]).

O objectivo desta tese é efectuar um estudo direccionado à avaliação da qualidade de experiência em serviços de videochamada, determinando quais os factores que a afectam e como representar essas eventuais dependências ou relações.

1.2 Motivação e Objectivos

Com a crescente utilização de aplicações que permitem efectuar videochamadas sobre a Internet cresce também o interesse de avaliar o desempenho destes serviços na perspectiva do utilizador. O serviço de videochamada, estando assente numa ligação à Internet, sofrerá do mesmo tipo de interferências que uma ligação deste tipo está sujeita (tais como atraso ou perda de pacotes). A videochamada estará dependente também da largura de banda que a sua ligação à Internet pode oferecer e as inerentes taxas de *upload* e *download*.

Um *Internet Service Provider* pode garantir um nível pré-acordado de qualidade de serviço ao utilizador, no entanto põe-se a questão se mesmo esta qualidade de serviço terá um impacto considerável na qualidade de experiência que um utilizador avalia. Estudando-se o impacto que diferentes condições de qualidade de serviço produzem na qualidade de experiência pode possibilitar uma melhor oferta de qualidade de experiência ao utilizador por parte do fornecedor de serviço o que poderá tornar-se num facto vital no que respeita à adopção do serviço.

A análise sobre o comportamento das aplicações mais populares que oferecem serviços de videochamada (Skype, Windows Live Messenger), do ponto de vista da experiência áudio/vídeo oferecida ao utilizador, poderá oferecer uma melhor visão sobre os factores que afectam a qual-

idade de experiência. De forma a efectuar esta análise são estudadas métricas de QoS e de QoE associadas ao serviço de videochamada, pretendendo nesta tese responder às seguintes questões:

1. Quais os requisitos de QoS associados a QoE e qual a sua relação?
2. Quais os parâmetros que têm efeito no QoE de uma videochamada?
3. Quais as condições limite que produzem diferentes níveis de QoE?

Estas questões são debatidas e respondidas nesta dissertação através de experimentação e colecta de dados e posterior análise. Todos os procedimentos seguidos e respectivos resultados são descritos em detalhe.

1.3 Metodologia de Estudo

A metodologia de estudo que se usa neste trabalho consiste numa primeira pesquisa bibliográfica dos trabalhos relevantes nesta área e métodos a usar na execução deste trabalho. Esta análise providencia informação essencial relativa ao estudo da avaliação de QoE em análise de vídeo e áudio (mais propriamente, a fala). Esta primeira abordagem é importante para o desenho e construção de uma arquitectura de testes que permita a colecta de dados que se necessita para realizar um estudo qualitativo. Após o estudo inicial segue-se uma parte experimental em que os testes necessários são realizados, sendo feita uma posterior análise de resultados de forma a se retirar as conclusões apropriadas.

1.4 Organização da Dissertação

No capítulo 2 reúne-se uma descrição das tecnologias relevantes a este trabalho bem como trabalhos anteriores importantes para o estudo que foi aqui desenvolvido.

O capítulo 3 contém uma abordagem mais aprofundada aos métodos de análise de vídeo que se decidiram usar no decorrer deste trabalho.

No capítulo 4 detalha-se a metodologia de testes e procedimentos implementados para a recolha dos dados necessários para o estudo.

O capítulo 5 apresenta os testes efectuados, os resultados obtidos e a sua análise.

CAPÍTULO 1. INTRODUÇÃO

No capítulo 6 exprimem-se as considerações finais referentes ao trabalho efectuado nesta tese, os resultados obtidos são discutidos e sugere-se o trabalho futuro.

Capítulo 2

Estado da Arte

De forma a melhor compreender o estudo feito nesta dissertação irá-se primeiro abordar a relação QoS/QoE. De seguida define-se com mais pormenor a qualidade de experiência, o seu significado, a sua importância e o papel que desempenha. Proceder-se-á por fim a apresentar alguns métodos de avaliação de qualidade áudio e vídeo relevantes nesta área.

2.1 De Qualidade de Serviço a Qualidade de Experiência

Quando o QoS foi definido pelo *International Telecommunication Union Standardization Sector*(ITU-T), foi definido como um conjunto de parâmetros que reflete a satisfação do utilizador com o serviço. Pode-se dizer que de certa forma esta definição se adequa agora ao propósito do QoE. O QoE pode ser visto como orientado ao utilizador e o QoS como orientado à rede e aos seus parâmetros de desempenho.

Enquanto que o QoE é importante para a manutenção de um cliente (clientes satisfeitos ficam com o seu fornecedor de serviços) os parâmetros de QoS são tipicamente mais fáceis de medir. Desta forma torna-se importante correlacionar os dois de forma a permitir um controlo de QoE baseado em medições de parâmetros de QoS.

O controlo de QoE é parte fundamental para o fornecedor de um serviço. Um utilizador ou mesmo uma aplicação tem a opção de mudar de serviço tendo em conta as condições que lhe permitem obter a melhor relação qualidade-preço. Desta maneira os parâmetros QoE podem ser vitais para sustentar tomadas de decisão que se esperam automáticas, bem como na escolha automática de uma rede.

As métricas de QoS são frequentemente encontradas em contratos de serviços estando associadas a um certo nível de desempenho, robustez e fiabilidade. Desta forma torna-se interessante estudar-se mecanismos de controlo auto-adaptativos para controlo de QoS tendo em conta a QoE experienciada pelo utilizador. Como mostrado na Figura 2.4 um ciclo de controlo de QoE/QoS começa com a qualidade de experiência percebida pelo utilizador final, expressos por um conjunto de requisitos qualitativos e/ou quantitativos QoE/QoS. Estes requisitos são normalmente traduzidos através de *Service Level Specifications* (SLS) acordados entre as duas partes envolvidas [1]. É de salientar que no mercado emergente vários serviços são oferecidos em regime melhor esforço, ou seja, sem qualquer garantia de serviço.



Figura 2.1: QoE/QoS management life-cycle [1]

Os serviços de voz, como VoIP, são tipicamente sujeitos a alguma forma de controlo de QoS, o que não é a norma para *web browsing* e muitos dos serviços multimédia existentes hoje. Pode-se debater que com o aumento gradual e contínuo que tem vindo a ser obtido em largura de banda que o QoS passa a ser menos crítico. Os mecanismos de controlo de QoS lidam com parâmetros como o jitter, parâmetros que podem levar a deficiências na comunicação tais como chamadas perdidas, *re-buffering*, falta de sincronização áudio-vídeo, vídeo "pixelizado" entre outros. Os consumidores de banda larga de hoje esperam conectividade a todo o tempo com grande eficiência e rapidez, tornando-se deste modo cada vez mais intolerantes a este tipo de falhas [15].

Na Figura 2.2 podemos ver o impacto qualitativo geral entre o QoS e QoE, i.e. a forma como a degradação de QoS afecta a QoE percebida.

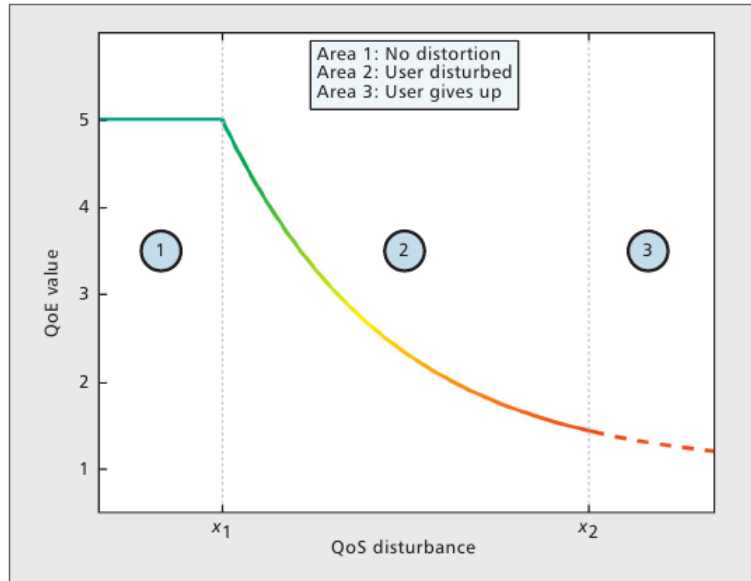


Figura 2.2: Forma geral da curva que mapeia a relação QoS/QoE [2]

2.2 Qualidade de Experiência

No contexto da interação homem-computador, a experiência de um utilizador (*user experience*) é definida como a percepção e resposta de um utilizador resultantes do uso ou antecipação do uso de um produto, sistema ou serviço. A experiência de um utilizador inclui todos os aspectos particulares a cada utilizador tais como emoções, crenças, preferências, percepções, reacções físicas e psicológicas, comportamentos e feitos que ocorrem antes, durante e depois do uso. Os três factores indicados que influenciam a experiência do utilizador são: o sistema, o utilizador e o contexto da utilização [16].

A avaliação da experiência do utilizador traduz o processo de investigação de como um utilizador avalia o uso de um produto, aplicação ou serviço. Devido à subjectividade da experiência do utilizador, que ainda é dependente do contexto e variável com o tempo, torna-se não-trivial avaliar esta de forma a obter resultados consistentes [17].

Existem vários métodos para avaliação de experiência do utilizador, não existindo uma única norma concreta de como proceder uma vez que existem vários aspectos a considerar. De acordo com os objectivos definidos neste trabalho irá-se focar no aspecto da avaliação geral da experiência do utilizador. Esta consiste numa estimação de como um utilizador avalia um sistema depois de o ter utilizado tempo suficiente. Estes métodos são no entanto adequados também a avaliar

episódios específicos ou momentâneos.

Como já referido, não existe somente um conjunto de procedimentos a tomar de forma a avaliar a experiência do utilizador devido às diferenças existentes entre as experiências que se pretende avaliar. Aponta-se no entanto um conjunto de objectivos gerais a atingir quando se pretende definir a experiência do utilizador:

- Utilidade: O utilizador achou as funções do sistema úteis e capazes de atingir o objectivo a que se propunham?
- Usabilidade: O utilizador acha que é fácil e eficiente utilizar o sistema?
- Estético: O utilizador gosta do aspecto visual do sistema?
- Identificação: O utilizador identifica-se com o produto?
- Estimulação: O sistema estimula o utilizador a criar por exemplo?
- Valor: O sistema é importante para o utilizador? qual a sua mais-valia?

Na Figura 2.3 apresenta-se uma representação dos factores que influenciam um utilizador.

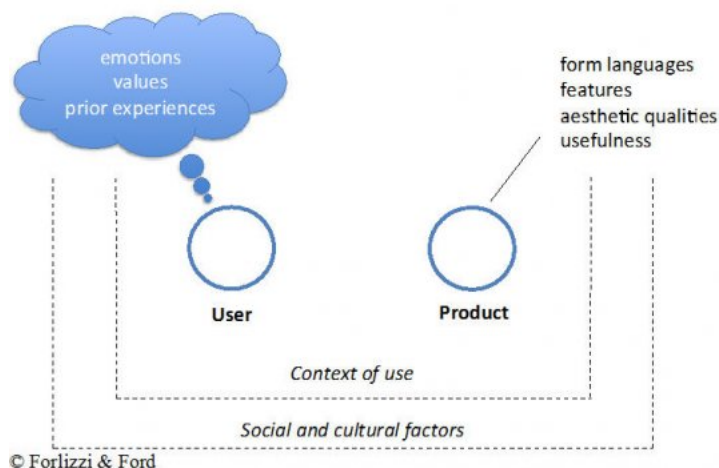
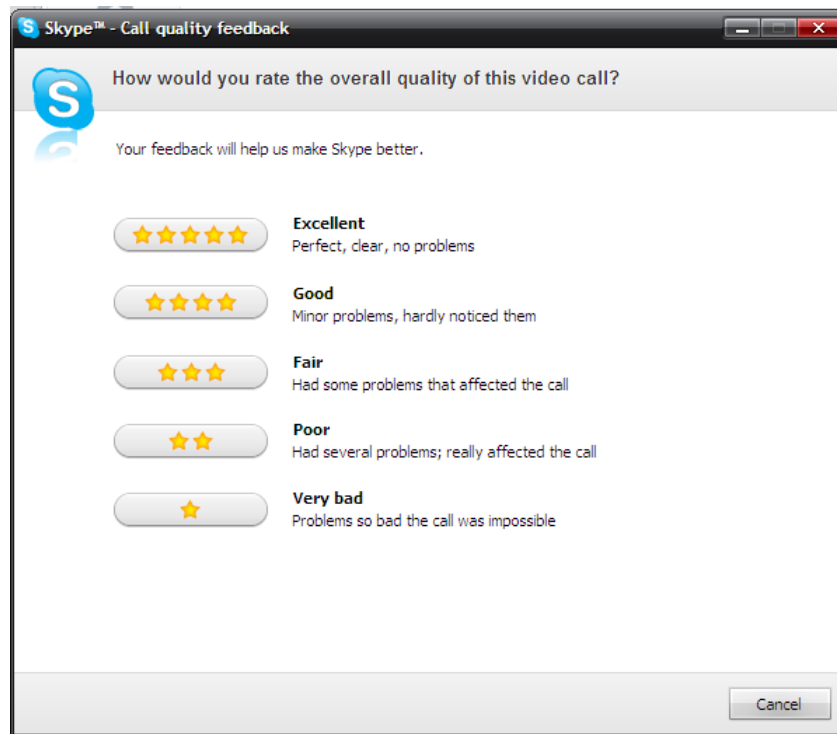


Figura 2.3: Influências na experiência do utilizador - Forlizzi & Ford [3]

Os métodos mais usados para recolher os dados que permitem efectuar uma avaliação são apresentados na forma de entrevistas, questionários ou uma combinação dos dois. Apresenta-se

de seguida o inquérito várias vezes apresentado no final de uma videochamada utilizando a aplicação Skype™ com o objectivo de avaliar a qualidade de experiência geral que o utilizador experimentou. Expõe-se primeiramente ao utilizador um questionário em que se pede uma avaliação baseada numa escala de *Mean Opinion Score* (MOS) (explicada em detalhe na secção seguinte), esta avaliação é exibida na Figura 2.4.



The image shows a window titled "Skype™ - Call quality feedback". Inside, there is a question: "How would you rate the overall quality of this video call?". Below the question is a subtext: "Your feedback will help us make Skype better." There are five rating options, each with a star icon and a description:

- Excellent**: Perfect, clear, no problems (5 stars)
- Good**: Minor problems, hardly noticed them (4 stars)
- Fair**: Had some problems that affected the call (3 stars)
- Poor**: Had several problems; really affected the call (2 stars)
- Very bad**: Problems so bad the call was impossible (1 star)

A "Cancel" button is located at the bottom right of the window.

Figura 2.4: Avaliação da qualidade geral da chamada

Após este inquérito inicial, caso a avaliação feita por parte do utilizador não seja excelente, este será exposto a um novo questionário, figura 2.5, de forma a identificar quais os problemas em concreto que afectaram a sua experiência durante a videochamada, é de notar também a separação entre as componentes áudio e vídeo.

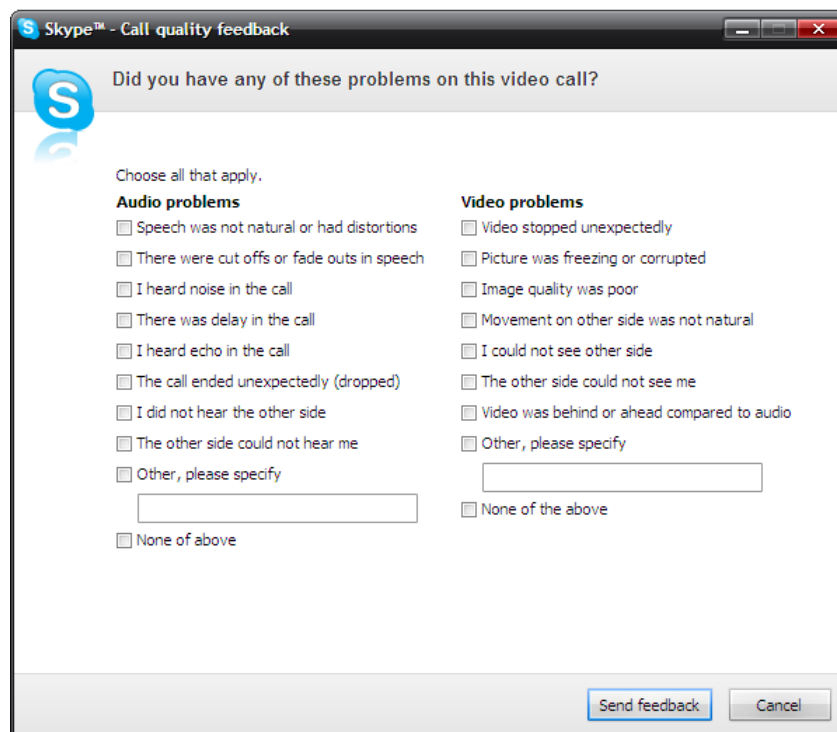


Figura 2.5: Identificação de problemas ocorridos

A ITU define Qualidade de Experiência como a aceitação geral de uma aplicação ou serviço, como percebida pelo utilizador final. Outra distinção entre QoS e QoE pode ser dita da seguinte maneira: O objectivo da rede e dos serviços deve ser atingir o máximo de avaliação por parte dos utilizadores (QoE), enquanto que a qualidade da rede (QoS) é a ferramenta principal para chegar a esse objectivo eficazmente [18].

Como medida de desempenho fim-a-fim no que toca a serviços do ponto de vista do utilizador final, a QoE é uma componente importante na arquitectura de sistemas. É particularmente relevante em serviços de vídeo onde um mau desempenho da rede pode afectar gravemente a experiência do utilizador. Quando desenhando sistemas deste tipo, a QoE é frequentemente pensada como um factor a tomar em conta.

Podemos dividir os métodos de avaliação da qualidade de experiência em dois tipos:

- Avaliação Subjectiva
- Avaliação Objectiva

Uma avaliação subjectiva consiste nas avaliações perceptuais dos utilizadores finais. Para

resultados sólidos devem-se repetir os testes consideravelmente de forma a se obter uma amostra aceitável. Para obter resultados concretos é necessário uma grande amostra de utilizadores o que torna o método mais dispendioso. Normalmente estes testes utilizam uma análise usando uma avaliação MOS [12][14] [19]. A avaliação MOS é detalhada na secção 2.3.

A avaliação Objectiva assenta em modelos matemáticos que estimam a avaliação que um utilizador humano faria. Este método muitas vezes preferido devido à rapidez e custo baixo que proporciona. Os resultados de uma avaliação objectiva devem ser similares aos de uma avaliação subjectiva. Nas secções 2.4, 2.5 e 2.6 serão detalhados alguns métodos de avaliação objectiva.

2.3 Mean Opinion Score

O Mean Opinion Score é um teste usado para obter a avaliação de uma rede ou serviço por parte do utilizador final. É portanto uma avaliação subjectiva. O MOS é especialmente usado na avaliação de serviços multimédia (voz, áudio ou vídeo). Este teste traduz a opinião dos utilizadores num valor numérico, de 1 a 5, onde 1 é o valor que indica a pior avaliação e 5 a melhor.

Tal como o nome indica, o MOS é obtido calculando a média dos resultados de um conjunto suficiente de avaliações. O MOS integra as recomendações do ITU [20]. A escala tipicamente usada nos testes é mostrada na Tabela 2.1.

MOS	Qualidade	Ruído
5	Excelente	Imperceptível
4	Bom	Perceptível mas não incómodo
3	Razoável	Ligeiramente incómodo
2	Pobre	Incómodo
1	Mau	Muito incómodo

Tabela 2.1: Mean Opinion Score

O MOS é um método de avaliação de qualidade de experiência em serviços multimédia extremamente sólido tendo como desvantagens o elevado custo e morosidade.

2.4 Peak Signal-to-Noise Ratio

Peak Signal-to-noise ratio (PSNR) designa o rácio entre o máximo possível da potência de um sinal e a potência do ruído que afecta a fidelidade da sua representação.

O PSNR é vulgarmente utilizado como uma medida da qualidade de reconstrução de um sinal (tal como um sinal de vídeo). O sinal é neste caso a representação original e o ruído é o erro introduzido pela sua reconstrução. O PSNR (método objectivo) pode ser usado como uma aproximação à avaliação de qualidade por parte de um humano, comparando reconstruções de sinais que foram afectados de formas diferentes ou até transmitidos através de meios distintos ou usando algoritmos de compressão diferentes. Numa comparação entre dois sinais um valor maior de PSNR indica uma maior aproximação ao sinal original, no entanto quando os sinais forem sujeitos a avaliação humana o valor PSNR poderá diferir da conclusão de maior qualidade da parte da avaliação humana [21].

O PSNR é obtido através do *mean squared error* (MSE), equação 2.1, em que para duas imagens $m \times n$ monocromáticas I e K , onde uma das imagens é considerada uma aproximação da outra devido a ruído, se tem

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (2.1)$$

sendo o PSNR definido como:

$$PSNR = 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE) \quad (2.2)$$

Na fórmula 2.2 MAX_I é o máximo valor possível do pixel da imagem. Usando PCM (*Pulse Code Modulation*) com B bits por amostra, MAX_I é $2^B - 1$.

Valores típicos de PSNR em compressão de imagem e vídeo encontram-se entre 30 e 50 dB [22]. Valores aceitáveis em transmissões sem fios situam-se entre os 20 a 25dB [23].

Quando os dois sinais são idênticos o valor do PSNR será indefinido. O PSNR é um algoritmo *full reference* o que significa que é necessário o sinal original e o reconstruído de forma a poder-se efectuar uma avaliação comparativa.

2.5 Perceptual Evaluation of Speech Quality

O *Perceptual Evaluation of Speech Quality* é um conjunto de normas composto por uma metodologia de teste para um avaliação automático semelhante ao de um utilizador humano para a qualidade de fala. O resultado é uma pontuação MOS. O PESQ é uma norma popularmente aplicada como referência para avaliação objectiva de discurso usado por fabricantes de telefones, de equipamento de rede e operadores de telecomunicações, estando definido na recomendação em [4].

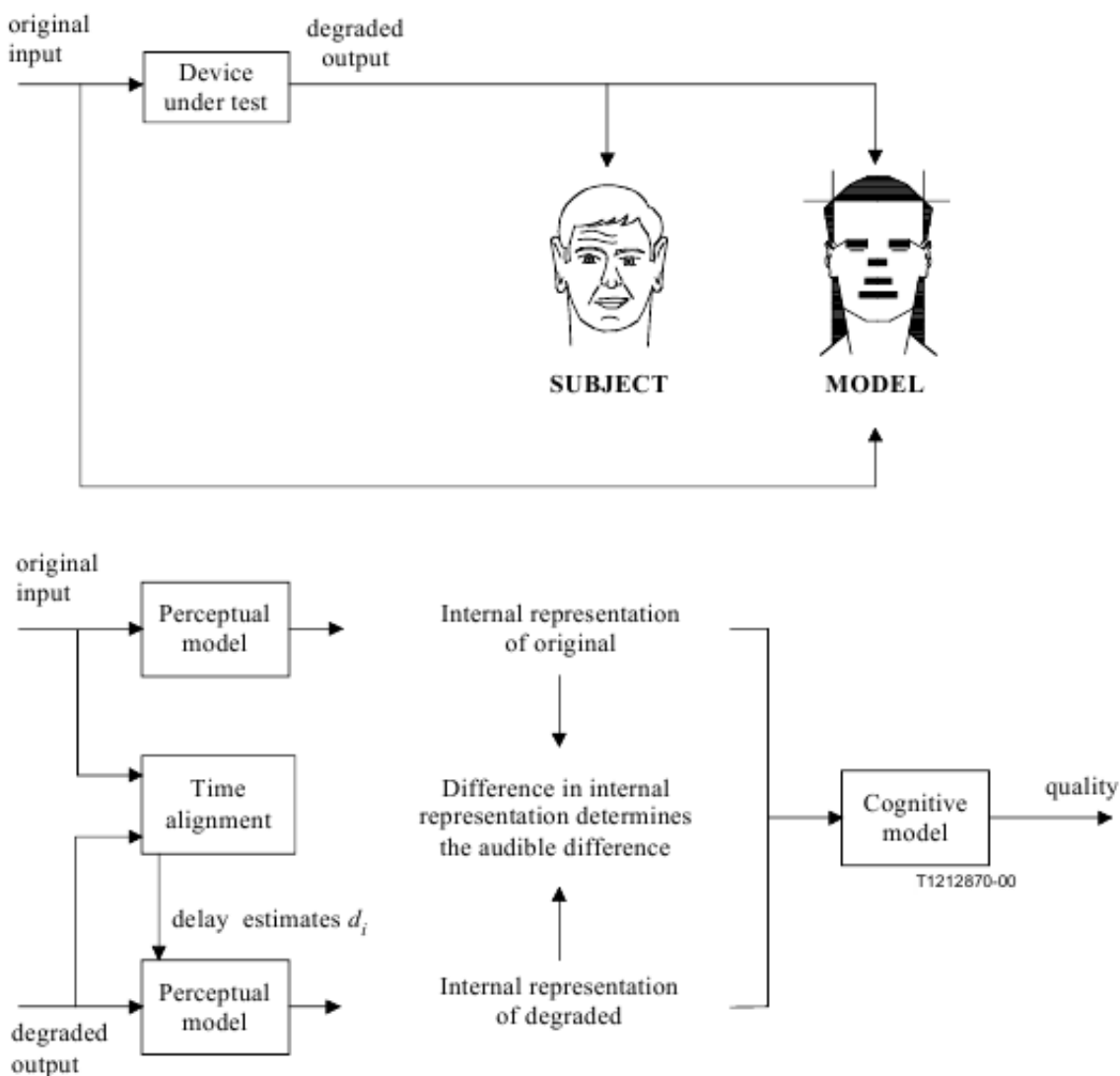
O PESQ foi desenvolvido com o objectivo de modelar testes subjectivos usados vulgarmente em telecomunicações para avaliar a qualidade de voz em humanos. O algoritmo utiliza vozes verdadeiras como sinais de teste. Para caracterizar a qualidade de audição percebida pelos utilizadores é imperativo usar sinais de fala em sistemas de telecomunicações. O conjunto de procedimentos para aplicações de amostras de voz estão definidos no guia PESQ ITU-T P.862.3.

O PESQ é um algoritmo *full reference* e analisa o sinal de voz (amostra a amostra) depois de um alinhamento temporal de fragmentos de referência e do sinal de teste.

O PESQ compara um sinal original $X(t)$ com um sinal degradado $Y(t)$ que é o resultado de $X(t)$ ter passado por um canal de comunicação. O resultado do PESQ é uma aproximação da qualidade que seria dada a $Y(t)$ por utilizadores num teste de audição subjectivo. No primeiro passo do PESQ uma série de atrasos entre o sinal original e degradado é calculada, um para cada intervalo de tempo para o qual o atraso é significativamente diferente do intervalo de tempo anterior. Para cada um destes intervalos, um ponto de começo e paragem correspondentes são calculados. O alinhamento do algoritmo é baseado no princípio de comparação da confiança de ter dois atrasos num determinado intervalo com a confiança de ter um único atraso nesse mesmo intervalo. O algoritmo consegue lidar com diferenças de atrasos tanto durante silêncios como durante períodos de fala.

Baseado no conjunto de atrasos que são identificados o PESQ compara o sinal original com o sinal degradado (já alinhado) do sistema sob teste usando um modelo perceptual, como ilustrado na Figura 2.6. O ponto fulcral neste processo é a transformação do sinal original e degradado para uma representação interna que é análoga à representação de sinais de áudio no aparelho auditivo humano. A representação interna é feita tendo em conta efeitos tais como variações de ganho e de filtragem linear, de modo que, caso estas variações não sejam muito acentuadas, tenham pouca significância na avaliação perceptual. Efeitos mais severos ou variações rápidas são apenas parcialmente compensadas e desta forma um efeito residual estará presente e contribuirá

para a avaliação perceptual, permitindo que um pequeno número de indicadores de qualidade sejam usados para modelar os resultados subjectivos. O PESQ utiliza dois parâmetros de erro no modelo cognitivo, estes são combinados e obtém-se uma avaliação objectiva MOS ([4] cap. 1).



NOTE – A computer model of the subject, consisting of a perceptual and a cognitive model, is used to compare the output of the device under test with the input, using alignment information as derived from the time signals in the time alignment module.

Figura 2.6: Esquema geral do funcionamento do PESQ ([4] Fig. 1)

As avaliações subjectivas são determinadas por diversos factores tais como as preferências individuais de cada utilizador e o contexto da experiência. Deste modo é necessário utilizar um

processo regressivo antes que uma comparação directa possa ser feita. Esta regressão deverá ser monotónica de modo que a informação seja preservada, e é normalmente usada para mapear o resultado objectivo PESQ com o resultado subjectivo. Uma boa avaliação de qualidade objectiva deve ter uma boa correlação com um grande número de experiências subjectivas se este processo regressivo é utilizado separadamente para cada experiência, na prática o mapeamento deste processo regressivo é habitualmente linear, usando uma escala similar ao MOS ([4] cap. 7).

O grau de semelhança entre o PESQ e os resultados subjectivos podem ser medidos calculando o coeficiente de correlação. Usualmente isto é realizado com as médias dos resultados, depois do mapeamento dos resultados objectivos com os subjectivos. Este coeficiente de correlação é calculado com a fórmula de *Pearson*:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (2.3)$$

Na fórmula 2.3, x_i é a condição MOS para a condição i , e \bar{x} é a média sobre os resultados MOS, y_i é a média da condição mapeada do resultado PESQ para a condição i , e \bar{y} é a média sobre os resultados MOS da condição prevista.

Para 22 experiências de *benchmark* ITU, a correlação média foi 0.935. Para um conjunto de 8 experiências usadas na validação final, a correlação média foi também 0.935 ([4] cap. 7.1).

2.6 NTIA General Model - VQM

Nesta secção irá-se detalhar o *NTIA General Model*. A decisão de se detalhar este modelo e não muitos outros prende-se com o facto de este modelo ser a base a partir da qual o modelo VQM - vfd, que foi o modelo escolhido para realizar os testes práticos no âmbito desta dissertação e descrito na secção 3.3, foi construído.

O *National Telecommunications and Information Administration* (NTIA) *General Model* é um método objectivo de avaliação de qualidade de vídeo. Foi construído de forma a ser um modelo de qualidade de vídeo (VQM) para uso geral em sistemas de vídeo de qualidade e bit rate variado. O *General Model* foi sujeito a vastos testes subjectivos e objectivos de maneira a verificar a sua desempenho antes de ser submetido ao teste VQEG Phase II [24]. Este modelo foi desenvolvido usando uma grande variedade de sistemas de vídeo e deve desta forma funcionar

correctamente com múltiplos tipos de codificação e sistemas de transmissão (ex. velocidades de transmissão de 10 kbits/s até 45 Mbits/s, MPEG-1/2/4, sistemas de transmissão digitais com erros e sistemas de transmissão analógicos). O *General Model* utiliza tecnologia de referência reduzida [25] e oferece estimativas das impressões sobre a qualidade de vídeo (i.e., Mean Opinion Score, como dado por um painel de utilizadores).

2.6.1 Técnicas de Calibração

O *General Model* e as suas técnicas de calibração associadas compreendem um sistema completo de avaliação objectiva de medição de qualidade de vídeo, como representado na Figura 2.7. A calibração dos vídeos originais e processados inclui alinhamento espacial, estimativa da região válida, compensação do ganho e nível, e alinhamento temporal. O cálculo VQM envolve a extracção de características baseadas na percepção, o cálculo de parâmetros de qualidade de vídeo, e combinar parâmetros de forma a construir o *General Model* ([5] cap. I).

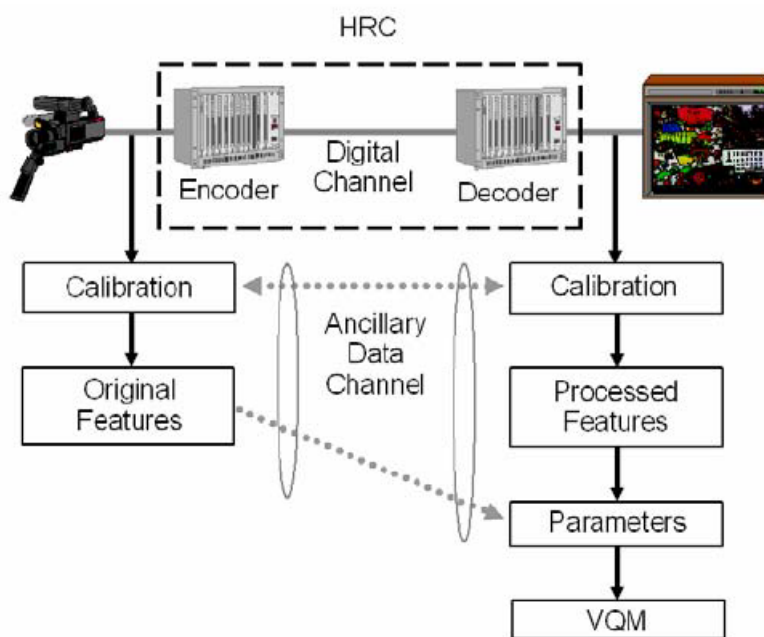


Figura 2.7: Diagrama de blocos do VQM ([5] Fig.1)

Os processos usados de forma a construir este modelo são os seguintes:

- **Alinhamento Espacial** - O alinhamento espacial determina o deslocamento espacial horizontal e vertical do vídeo processado em relação ao vídeo original. A precisão do algoritmo

de alinhamento espacial vai até ao 0.5 pixel mais próximo em deslocamentos horizontais e até à linha mais próxima em deslocamentos verticais. Depois do alinhamento espacial ter sido calculado, o deslocamento espacial é removido do vídeo processado. O alinhamento espacial deve ser determinado antes da região válida processada, abreviada como PVR (*processed valid region*), que define a porção da imagem processada que contém informação válida, ganho e compensação de nível, e alinhamento temporal. O algoritmo de alinhamento espacial calcula o alinhamento espacial para cada série de *frames* processados a uma frequência determinada, tal como 1 *frame* a cada segundo. Usando esta estimativa como ponto de partida, o algoritmo realiza pesquisas minuciosas de forma a encontrar o *frame* original que mais se assemelha ao *frame* processado, e realiza também estimações de ganho e compensação de nível. Para alguns *frames* processados, o alinhamento espacial pode falhar. Normalmente quando este alinhamento incorrecto acontece para um *frame* processado, a ambiguidade deve-se a características da cena ([5] cap. II).

- **Região Válida Processada** - As amostras de vídeo podem ter uma fronteira de pixels e linhas que não contém uma imagem válida. Um sistema de vídeo digital que utiliza compressão pode reduzir ainda mais a área da imagem de forma a diminuir o número de bits enviados numa transmissão. Para evitar que estas áreas da imagem que não contém informação válida influenciem os resultados VQM, estas áreas devem ser excluídas do cálculo VQM. Como o comportamento de alguns sistemas de vídeo é dependente da cena presente, a região válida deve ser determinada usando as sequências de vídeo reais. Neste caso, a PVR deve ser calculada para cada cena separadamente. Após a PVR ter sido calculada, os pixels inválidos são descartados tanto da sequência de vídeo original como da processada. O algoritmo começa com a suposição que as margens exteriores de cada *frame* processado contém informação inválida. A dimensão desta região inválida é estabelecida de forma empírica, baseada nas observações de sistemas de vídeo reais. Os pixels imediatamente dentro da actual estimativa da região válida são nessa altura examinados. Se o valor médio dos pixels é preto ou um valor próximo de preto, a estimativa da região válida é diminuída em tamanho. As condições de paragem podem ser atingidas incorrectamente devido ao conteúdo da imagem. Por exemplo, se uma imagem contém preto do lado esquerdo (que faz parte da cena) isto irá causar com que o algoritmo pressuponha que a coluna válida que se situa mais à esquerda da imagem estará situada mais ao meio da imagem do que deveria estar. Por essa razão, o algoritmo é aplicado a múltiplas imagens do vídeo processado e a maior PVR observada (com margem de segurança) é usada para a estimativa de PVR final ([5] cap. III).

- **Ganho e Compensação de Nível** - De forma a efectuar a calibração de ganho e compensação de nível é necessário que as imagens originais e processadas tenham sido alinhadas espacialmente e temporalmente. O método usado neste modelo assume que os sinais Y , C_B , e C_R da Rec. 601 têm um ganho e compensação de nível independente. Esta suposição é normalmente suficiente de forma a calibrar sistemas de vídeo (ex. Y , $R-Y$, $B-Y$). As regiões válidas dos *frames* originais e processados são divididos em pequenas sub-regiões quadradas, ou blocos. A média sobre o espaço das amostras $[Y, C_B, C_R]$ para cada sub-região correspondente original e processada é calculada para formar sub-amostras de imagens. Para registar temporalmente um *frame* processado (com deslocamento espacial constante), o desvio padrão de cada (original menos processado) diferença de imagem é calculado usando os *frames* de luminância Y sub-amostrados. Para um dado *frame* processado, o desvio temporal que produz o menor desvio padrão é escolhido como a melhor correspondência. Um ajuste linear de primeira ordem é usado para calcular o ganho relativo e a compensação entre os *frames* originais e processados sub-amostrados. Este ajuste linear é aplicado independentemente a cada um dos três canais: Y , C_B , e C_R ([5] cap. IV).
- **Alinhamento Temporal** - Atrasos excessivos de vídeo impedem uma comunicação bidireccional eficaz. O atraso no vídeo pode depender dos atributos dinâmicos da cena da sequência de vídeo (ex. detalhe espacial, movimento) e do sistema de vídeo (ex. débito). Um vídeo com grande quantidade de movimento pode sofrer mais atraso que vídeos de pouco movimento. Estimativas de atraso do vídeo são necessárias para alinhar temporalmente o vídeo original e processado antes de se proceder à medição de qualidade. A técnica usada para estimar o atraso do vídeo baseia-se nos *frames*, no sentido em que funciona correlacionando imagens de resolução baixa, sub-amostradas no espaço e extraídas das sequências de vídeo originais e processadas. Esta técnica "*frame-based*" estima o atraso de cada *frame*. Estas estimativas individuais são combinadas para estimar o atraso médio da sequência de vídeo. Para reduzir a influência ou distorções no alinhamento temporal, as imagens originais e processadas são espacialmente sub-amostradas e depois normalizadas de forma a terem variância unitária. Cada imagem individual processada é então temporalmente registada usando a técnica usada no algoritmo do ganho e compensação de nível (i.e. encontra a imagem original que minimiza o desvio padrão da diferença entre as imagens originais e processadas). Isto identifica a imagem original mais semelhante para cada imagem processada. No entanto, não é a identidade da imagem original que é de interesse, mas sim o atraso relativo entre a imagem original e processada (em segundos ou *frames*). Ao contrário dos algoritmos de calibração prévios, o algoritmo de calibração

temporal examina cada *frame* processado ([5] cap. V).

2.6.2 Características e Parâmetros do General Model

Uma característica de qualidade no contexto deste algoritmo é definida como a quantidade de informação associada com, ou extraída de, uma sub-região espacial-temporal de uma sequência de vídeo (original ou processada). Estas características são uma função de espaço e tempo que caracterizam mudanças perceptuais nas propriedades espaciais, temporais e de cromaticidade das sequências de vídeo. Ao comparar características extraídas do vídeo processado calibrado com características extraídas do vídeo original, parâmetros de qualidade podem ser calculados que são indicativos de mudanças perceptuais na qualidade de vídeo. Conceptualmente, todas as características usadas no *General Model* executam os mesmos passos. Um filtro perceptual é aplicado à sequência de vídeo para melhorar alguma propriedade da qualidade de vídeo percebida. Depois desta filtragem perceptual, características são extraídas de sub-regiões espaciais-temporais (S-T, *spatial-temporal*) usando uma fórmula matemática (ex. desvio padrão). Finalmente um limite de perceptibilidade é aplicado às características extraídas. Cada filtro perceptual distingue um aspecto de qualidade de vídeo ([5] cap. VI).

- **Parâmetros do *General Model*** - O *General Model* contém sete parâmetros independentes. Quatro parâmetros são baseados em características extraídas de gradientes espaciais do componente de luminância Y, dois parâmetros são baseados em características extraídas do vector formado pelos dois componentes de cromaticidade (C_B , C_R), e um parâmetro é baseado no produto de características que medem contraste e movimento, as quais são ambas extraídas do componente de luminância Y ([5] cap. VII). Os sete parâmetros são:

1. Parâmetro "si_loss"

Este parâmetro detecta um decréscimo ou perda de informação espacial (ex. desfocagem)

2. Parâmetro "hv_loss"

O parâmetro *hv_loss* detecta um deslocamento de margens de orientações horizontais e verticais para orientações diagonais, como pode ser o caso se margens horizontais e verticais sofram mais esbatimento que margens diagonais.

3. Parâmetro "hv_gain"

Este parâmetro detecta um deslocamento de margens da diagonal para a horizontal e vertical.

4. Parâmetro "chroma_spread"

Este parâmetro detecta mudanças na distribuição de amostras de cor bidimensionais.

5. Parâmetro "si_gain"

Este é o único parâmetro de melhoramento de qualidade presente no modelo. Este parâmetro mede melhoramentos na qualidade que resultam do aperfeiçoamento das margens.

6. Parâmetro "ct_ati_gain"

A perceptibilidade de deficiências espaciais pode ser influenciada pela quantidade de movimento existente. Da mesma forma, a perceptibilidade de deficiências temporais pode ser influenciada pela quantidade de detalhe espacial presente. Uma característica derivada do produto da informação de contraste e informação temporal pode ser usada para, em parte, explicar estas interações. A métrica *ct_ati_gain* é calculada como o produto de uma característica de contraste, medindo a quantidade de detalhe espacial, e uma característica de detalhe temporal, medindo a quantidade de movimento na região S-T. As falhas serão mais visíveis nas regiões S-T que têm um produto baixo que nas regiões S-T com um produto alto. Isto será mais evidente em deficiências tais como ruído e blocos de erro. O parâmetro *ct_ati_gain* identifica falhas de flutuação de margens que estão quase sempre presentes.

7. Parâmetro "chroma_extreme"

Esta característica usa as mesmas características de cor que a métrica *chroma_spread*, mas diferentes funções de projecção espaciais e temporais. *chroma_extreme* detecta graves deficiências de cor, tais como as produzidas por erros de transmissão digital.

2.6.3 Cálculo e Desempenho do General Model

Esta secção descreve como o *General Model* é calculado usando os valores dos parâmetros. O *General Model* é optimizado de forma a alcançar a máxima correlação objectiva/subjectiva usando uma grande variedade de qualidade de vídeo e bit rate. O *General Model* produz resultados que vão desde zero (sem falhas perceptíveis) até aproximadamente um (máxima insuficiência percebida). Os valores do *General Model* podem ser multiplicados por 100 de forma a aproximar os resultados a uma escala DMOS (*Difference*

Mean Opinion Score) derivada da escala DSCQS (*Double Stimulus Continuous Quality Scale*). O *General Video Quality Model* (VQM) consiste na seguinte combinação linear dos sete parâmetros dados no ponto anterior:

$$\begin{aligned} VQM = & -0.2097 \times si_loss + 0.5969 \times hv_loss + 0.2483 \\ & \times hv_gain + 0.0192 \times chroma_spread - 2.3416 \times si_gain \\ & + 0.0431 \times ct_ati_gain + 0.0076 \times chroma_extreme \end{aligned} \tag{2.4}$$

O parâmetro *si_loss* é sempre menor ou igual a zero, desta forma este parâmetro apenas pode aumentar o VQM. Como todos os outros parâmetros são maiores ou iguais a zero, *si_gain* é o único parâmetro que pode diminuir o VQM ([5] cap. VIII).

O propósito fundamental do *General Model* e as rotinas de calibração associadas é fornecer resultados de qualidade vídeo subjectivos. Este propósito foi fundamentado comparando os seus resultados com avaliações subjectivas de vídeos. O *General Model* foi desenvolvido usando dados de teste subjectivos e objectivos de onze experiências diferentes sobre qualidade de vídeo. Todos estes testes foram conduzidos de acordo com a recomendação ITU-R BT.500 [26] ou a recomendação ITU-T P.910 [27]. Todas as experiências foram conduzidas usando sequências de vídeo de 8 a 10 segundos de duração.

Juntando todas as experiências, estas incluem 1536 sequências de vídeo avaliadas subjectivamente. A Figura 2.8 mostra o gráfico de dispersão da qualidade subjectiva em relação ao VQM, onde a cada sequência de vídeo foi dada uma cor diferente. O coeficiente de correlação linear Pearson geral entre a qualidade subjectiva e o VQM para as sequências de vídeo é ilustrado na Figura 2.8 é de 0.948. A Figura 2.9 mostra a média dos resultados por cenas de vídeo de forma a produzir um único resultado subjectivo (i.e. qualidade subjectiva HRC) e resultado objectivo (i.e. HRC VQM) para cada sistema de vídeo. Qualidade subjectiva HRC é indicativa de como o sistema responde (em média) a um conjunto de cenas de vídeo. O coeficiente de correlação linear Pearson geral entre a qualidade subjectiva HRC e HRC VQM para os dados na Figura 2.9 é de 0.980.

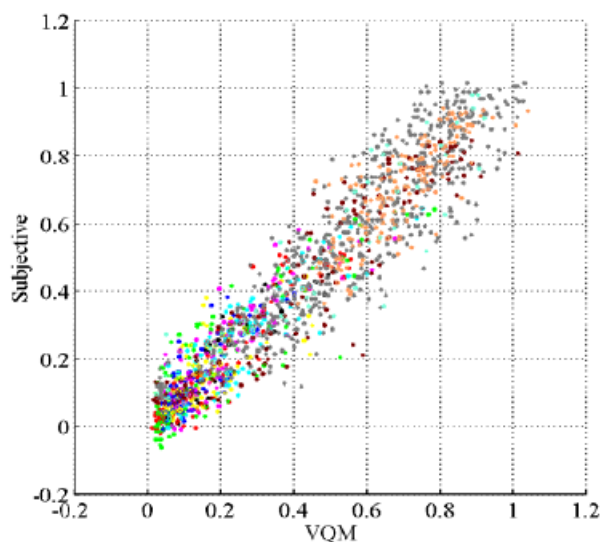


Figura 2.8: Qualidade subjectiva de vídeo vs VQM ([5] Fig. 2)

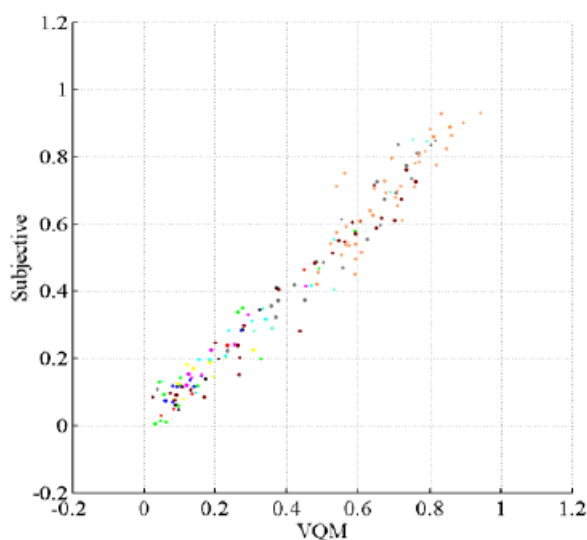


Figura 2.9: Qualidade subjectiva de vídeo HRC vs HRC VQM ([5] Fig. 3)

O VQEG (*Video Quality Experts Group*) oferece *input* aos organismos de normalização responsáveis pela produção de recomendações relativas a métricas objectivas de qualidade de vídeo. Com este propósito o VQEG realizou o teste FR-TV Phase II, de 2001 a 2003 [24]. Estes testes ofereceram uma avaliação independente da capacidade de modelos de qualidade de vídeo e os seus algoritmos de calibração associados para reproduzir resultados subjectivos. Este teste conteve duas experiências, uma para vídeo de 525 linhas e outra para vídeo de 625 linhas. Os

testes subjectivos foram realizados em 3 laboratórios independentes.

Os dados subjectivos e VQM para estas duas experiências estão representadas nas Figuras 2.10 e 2.11. No teste de vídeo de 525 linhas, o *General Model* foi um de dois modelos que estatisticamente cumpriu o objectivo melhor que todos os outros modelos testados. A correlação linear de Pearson foi de 0.938. No teste de 625 linhas o *General Model* foi um de quatro modelos que estatisticamente cumpriu o objectivo melhor que todos os outros modelos testados. A correlação linear de Pearson foi de 0.886.

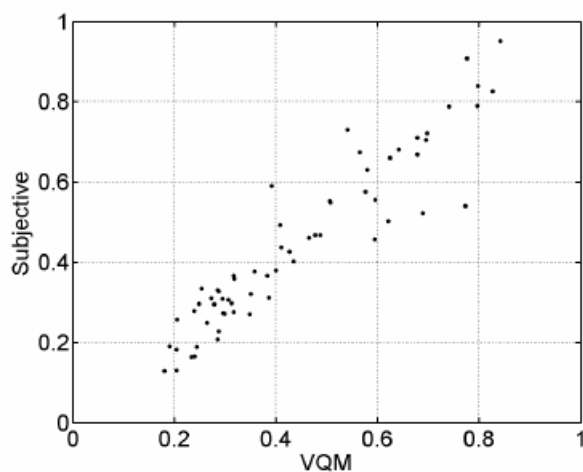


Figura 2.10: VQEG FR-TV Phase II - 525 linhas: Qualidade subjectiva de vídeo vs VQM ([5] Fig. 4)

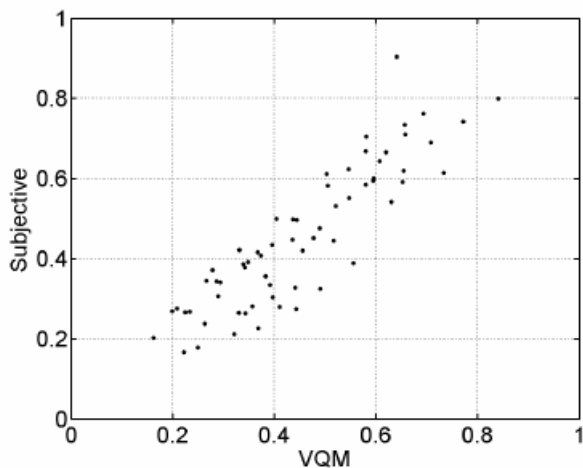


Figura 2.11: VQEG FR-TV Phase II - 625 linhas: Qualidade subjectiva de vídeo vs VQM ([5] Fig. 5)

Globalmente, o VQM foi o único modelo que estatisticamente melhor cumpriu o objectivo em ambos os testes de 525 linhas e 625 linhas. Obteve um coeficiente de correlação Pearson médio em ambos os testes de 0.91, o VQM foi o único modelo a ultrapassar o limite de 0.9. Como resultado, o VQM foi normalizado pela ANSI em Julho de 2003 (ANSI T1.801.03-2003), e foi incluído em Draft Recommendations do ITU-T Study Group 9 e ITU-R Working Party 6Q ([5] cap. IX).

2.7 Resumo

Neste capítulo foram primeiramente abordados os conceitos de QoE e QoS e a sua relação. De seguida introduziram-se os conceitos de avaliação subjectiva e objectiva de QoE distinguindo o que define cada uma. Finalmente referiram-se exemplos concretos de métodos de avaliações subjectivas e avaliações objectivas.

No próximo capítulo, Capítulo 3, passar-se-á a pormenorizar e fundamentar os métodos que foram escolhidos para se efectuarem os testes práticos de análise de vídeo realizados neste trabalho.

Capítulo 3

Métodos para Análise de Vídeo

Neste capítulo começa-se por introduzir primeiro a ferramenta para análise de vídeo denominada BVQM e as opções de avaliação que este disponibiliza. De seguida oferece-se uma descrição mais detalhada do método de calibração e modelo de avaliação usados ao longo deste trabalho para se efectuar a análise de vídeo.

3.1 Batch Video Quality Metric

A NTIA desenvolveu uma ferramenta de software de medição de qualidade de vídeo (BVQM) que realiza a análise de múltiplas sequências de vídeo para determinar objectivamente a sua qualidade de vídeo. O BVQM possui opções de medição de qualidade tais como:

- realizar a calibração das sequências de vídeo (i.e. dimensionamento espacial, registo espacial, estimativa da região válida, compensação de ganho e nível, registo temporal);
- determinar quais os modelos de qualidade de vídeo a usar para uso na estimativa da percepção global de qualidade.

O BVQM tem um interface gráfico, representado na Figura 3.1, para seleccionar as sequências de vídeo que se desejam analisar. A análise consiste na calibração dos cliques de vídeo processados e calcular as suas métricas de qualidade de vídeo associadas. É possível estabelecer valores de calibração manualmente ou escolher entre várias opções de calibração automáticas. Múltiplos modelos de qualidade vídeo estão também disponíveis. Depois do processamento, o

BVQM exhibe os resultados graficamente e proporciona relatórios em ficheiros de texto. Todos os resultados ficam disponíveis para visualização ou para exportação para folhas de cálculo.

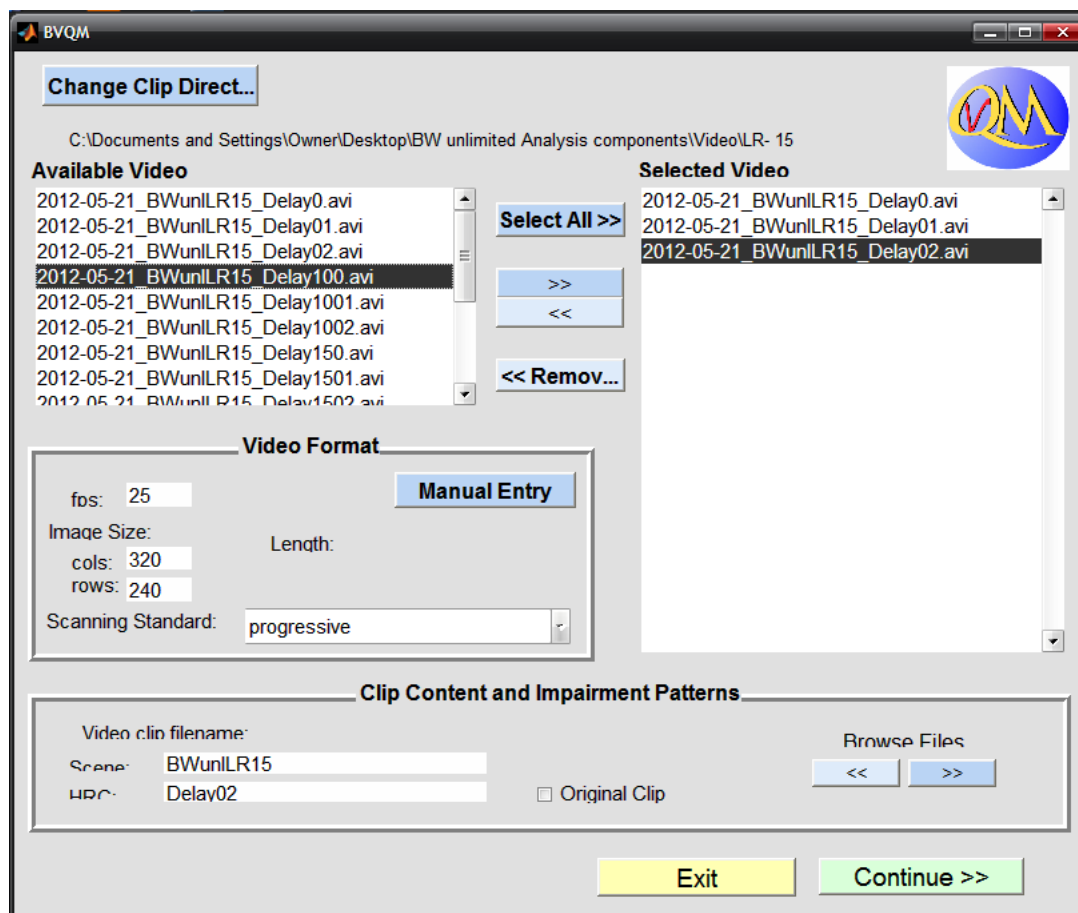


Figura 3.1: Interface gráfico BVQM

O BVQM apenas pode trabalhar com ficheiros de vídeo de certos formatos. O primeiro formato suportado pelo BVQM é o *Audio Video Interleave* (AVI). Apenas ficheiros AVI não comprimidos são suportados. Estes ficheiros AVI não comprimidos devem conter o formato UYVY, YUY2, YV12, V210, ou RGB24. O BVQM opera mais rapidamente em ficheiros AVI que contenham um formato que não seja RGB24. Ficheiros AVI comprimidos devem ser convertidos primeiramente em ficheiros descomprimidos antes da sua análise. Note-se que distorções causadas pela compressão dos ficheiros podem influenciar os resultados ([28] cap. 1).

O BVQM automaticamente determina como processar os ficheiros de vídeo examinando o nome destes ficheiros. O tempo de configuração é bastante menor quando nomes normalizados são usados para todos os ficheiros. A convenção para o nome dos ficheiros a analisar pelo BVQM

é "cena_hrc" seguido do sufixo ".avi" para ficheiros AVI ou ".yuv" para ficheiros Big-YUV. Os nomes "cena" e "hrc" contêm apenas letras e números. O utilizador pode especificar uma convenção diferente de nomes na forma de "teste_cena_hrc.avi" ou "teste_cena_hrc.yuv". Contudo, a versão actual do BVQM pode apenas analisar um vídeo de cada vez (i.e. uma lista de vídeos da mesma experiência ou teste).

"Cena" identifica uma sequência, ou segmento, de vídeo. Todas as versões da mesma cena devem ter o mesmo nome. "HRC" (*Hypothetical Reference Circuit*) indica um sistema de vídeo sob teste tal como uma codificação a operar a um determinado bit rate ou um canal de transmissão digital. Quando um parâmetro é modificado na codificação, no canal de transmissão, ou na decodificação, deverá ser dado um diferente HRC. O BVQM apresenta maior precisão quando examina múltiplas cenas que foram passadas pelo mesmo HRC. O nome HRC "original" refere-se à versão da cena de maior qualidade disponível, e esta será usada como a sequência de referência para as medições dessa cena. O original é normalmente a sequência de vídeo que foi passada pelo HRC para criar as sequências de vídeo processadas. Preferencialmente o vídeo original não é comprimido. As sequências de vídeo originais devem de ser excelente ou boa qualidade quando examinadas visualmente. As sequências de vídeo originais devem estar disponíveis para o BVQM ([28] cap. 4.2).

A calibração tem como objectivo tentar remover imperfeições do vídeo processado que os utilizadores não vêem ou que não têm importância para o utilizador. Estas imperfeições incluem, por exemplo, o alinhamento espacial da sequência de vídeo uns poucos pixels para a direita. Se um HRC induziu tal deslocamento, e não é removido antes do cálculo do modelo, a estimativa de qualidade do modelo será significativamente afectada.

Os algoritmos de calibração *Full Reference* e *Reduced Reference* tornam-se mais efectivos quando o BVQM é utilizado com múltiplas cenas associadas ao mesmo HRC. Alguns parâmetros de calibração são sistemáticos, isto é, constantes para cada HRC, o que resulta em maior precisão analisando todas as cenas relacionadas com o mesmo HRC ([28] cap. 6.1). O BVQM apresenta as seguintes opções de calibração:

- Sem calibração - Não efectua qualquer tipo de calibração. Assume que o vídeo original e o processado estão perfeitamente alinhados;
- Calibração manual - Lê os valores de calibração de uma folha de cálculo modificada pelo utilizador;
- *Temporal Registration after Manual Calibration* - Lê os valores de calibração de uma folha

de cálculo e de seguida efectua apenas o registo temporal;

- *Temporal Registration and Valid Region Only* - Efectua apenas o registo temporal e os algoritmos de estimativa da região válida;
- *Reduced Reference Calibration Version 1* - Realiza calibração *reduced reference* como explicado em [29]. Estes algoritmos usam processos aleatórios o que pode resultar em resultados diferentes em cada processo. Os algoritmos de referência reduzida podem exibir maus resultados quando a cena apresenta grande movimento;
- *Reduced Reference Calibration Version 2* - Calibração do tipo *reduced reference* como explicado em [30] [31]. A versão 2 é uma pequena melhoria da versão 1;
- *Full Reference Calibration* - Calibração *Full Reference* como dado em ANSI T1.801.03-2003, ITU-T J.144 (03/04), e ITU-R BT.1683 (06/04). Ao utilizador será pedido um valor a ser usado como incerteza temporal. O valor default é um segundo;
- *Peak Signal to Noise Ratio* - Executa uma calibração PSNR *full reference* como mostrado em ITU-T J.340;
- *ITU-T J.244 seguido de ITU-T J.340* - Executa uma calibração *reduced reference version 2* seguida da calibração PSNR. A ideia por detrás desta calibração é realizar a calibração PSNR para sequências com grande incerteza temporal e espacial mas sem longos tempos de espera;
- *ITU-T J.144 seguido de ITU-T J.340* - Realiza uma calibração *full reference*, seguida de uma calibração PSNR. A ideia desta calibração foi, também, de executar uma calibração PSNR sem longos tempos de execução;

Os modelos de qualidade vídeo estimam a média das opiniões de um grupo sobre a qualidade geral que estes experienciam quando observam uma sequência de vídeo. As opções de modelos que o BVQM apresenta são as seguintes:

- *NTIA General Model* - O *General Model* está normalizado pelo ANSI e incluído em duas recomendações ITU. Foi previamente descrito na secção 2.6;
- *Low Bandwidth Model* - Este modelo utiliza características *reduced reference* que minimizam a necessidade de largura de banda para transmitir a informação necessária. A velocidade de processamento deste modelo é semelhante à do *General Model*. Está totalmente documentado em [32];

- *Developer's Model* - Minimiza o uso do CPU e é uma aproximação ao modelo NTIA. No entanto é mais rápido aproximadamente seis vezes que o *General Model*. Está documentado em [33];
- *Television Model* - Este modelo foi otimizado para uma maior qualidade de vídeo usualmente encontrado em aplicações de televisão. Documentação em [33];
- *Video Conferencing Model* - Otimizado para baixa qualidade de vídeo tipicamente encontrada em aplicações de vídeo conferência. Encontra-se documentado em [33];
- *Fast Low Bandwidth Model* - Este modelo minimiza a utilização do CPU. É uma aproximação do *Low Bandwidth Model*, mas quatro vezes mais rápido que este. Foi incluído na recomendação ITU-T J.249;
- PSNR - Este modelo calcula o PSNR usando um nível de sinal de 235 ou 255;
- *PSNR com Variable Frame Delay* - O PSNR_VFD calcula o PSNR depois de ter sido feita a correspondência usando VFD. Está documentado em [34];
- *VQM com Variable Frame Delay* - Esta opção selecciona o modelo VQM_VFD. Este modelo inclui estimativas e remoções VFD, parâmetros VFD que quantificam distorções perceptuais devido ao VFD, distancia de visualização e uma rede neural para executar o mapeamento dos valores dos parâmetros para resultados subjectivos MOS. Documentação completa em [7] e descrita na secção 3.3.2.

3.2 Método *Full Reference* para estimar atrasos de vídeo

Sistemas de transmissão digitais constituídos por um codificador de vídeo, um método de comunicação digital (ex. Internet Protocol - IP), e um decodificador de vídeo podem produzir pausas na apresentação do vídeo, após as quais o vídeo pode prosseguir saltando ou não frames de vídeo. O sistema pode por exemplo decidir baixar o número de frames por segundo da transmissão para diminuir o número de bits a transmitir. Outra razão pode ser que o sistema tenha decidido pausar o último frame correcto que obteve quando a transmissão digital é interrompida quando erros de perda de pacotes ocorrem. Esta variação de tempo dos frames de vídeo processado constituem um desafio para alguns sistemas de avaliação de qualidade de vídeo. A razão é que os erros de alinhamento temporal entre as sequências de vídeo originais e processadas podem

produzir erros de avaliação para medições *full reference* tais como PSNR que excedem em muito o impacto perceptual da variação dos atrasos de tempo em vídeos. Por exemplo, uma pausa de um frame num vídeo, sem saltar frames, pode resultar no segmento processado anterior ou posterior ser deslocado em um frame em relação ao segmento original. Enquanto isto é quase não notado pelos utilizadores, a medição PSNR irá detectar uma grande falha no segmento processado que está deslocado por um frame em relação ao original.

O método *full reference* é um método para estimar atrasos de vídeo variáveis. A melhor correspondência de frame de vídeo original é determinada para cada frame de vídeo do vídeo processado. Para algumas sequências de vídeo, estimar a melhor correspondência de frames de vídeo originais é difícil. Isto é verdade para vídeo que não contenha movimento ou próximo disso, vídeo com movimento repetitivo, vídeo processado bastante distorcido, ou vídeo processado com actualizações parciais de frames. Para estas sequências difíceis, um processamento causal sofisticado (i.e. o vídeo não pode voltar atrás no tempo) que junta a sequência alinhada do vídeo usando um conjunto de regras heurísticas pode melhorar as estimativas dos atrasos de vídeo. Este algoritmo pode também produzir uma lista de vídeo frames originais para cada frame de vídeo processado ordenada por possível grau de correspondência.

Uma possível aplicação para os resultados deste algoritmo é remover os atrasos variáveis de frames da sequência de vídeo processada como um processo de calibração antes de calcular uma medição *full reference* da qualidade de vídeo. Uma segunda aplicação é calcular um parâmetro de qualidade de vídeo que apenas mede os efeitos perceptuais dos atrasos variáveis de frames. Combinando este parâmetro de qualidade de vídeo com a medição de qualidade *full reference* produzirá uma estimativa geral da qualidade de vídeo mais precisa ([34] cap. 1).

O algoritmo *full reference* para determinar a melhor correspondência entre frames do vídeo original e processado foi desenvolvido usando sequências de vídeo que variam entre 8 e 15 segundos de duração. O comportamento do algoritmo para sequências de vídeo mais curtas ou mais longas deve ser analisado antes de ser aplicado. Esta secção oferece uma visão geral da descrição do algoritmo tal como ela é aplicada a um clipe de vídeo. O algoritmo utiliza apenas as imagens de luminância do vídeo.

O algoritmo completo que inclui o processamento causal pode ser dividido em 6 etapas. A primeira etapa ([34] cap. 2.1) envolve o cálculo dos MSE (*mean squared error*) entre cada frame processado e o conjunto de frames originais que estão compreendidos num intervalo temporal indicada pelo utilizador. Este passo computacionalmente pesado produz informação usada pelo resto do algoritmo (etapas 2 a 6). A segunda etapa ([34] cap. 2.2) realiza um procedimento de

limitação nos MSEs originais para cada frame processado de forma a produzir uma lista ordenada de alinhamentos de tempo para cada frame processado (ordenado de maior compatibilidade para menor compatibilidade). A etapa 3 ([34] cap. 2.3) examina a lista dos alinhamentos MSE menores para determinar segmentos de tempo que demonstram uma causalidade "normal", onde a causalidade normal impõe um limite definido pelo utilizador na magnitude de saltos de tempo de frame para frame. A etapa 4 (descrita em [34] cap. 2.4) começa a geração do alinhamento de tempo causal para o vídeo processado começando com o maior segmento causal normal (etapa 3) e preenchendo os segmentos causais normais mais pequenos. A etapa 5 ([34] cap. 2.5) preenche os restantes lugares usando interpolação ou a lista ordenada de MSEs gerada na etapa 2 (que inclui uma lista de alinhamentos temporais prováveis para cada frame processado em adição ao alinhamento temporal MSE mínimo mais provável). No fim da etapa 5, a melhor estimativa do alinhamento temporal causal é obtida. Na etapa 6, descrita em ([34] cap. 2.6), é expandida a lista ordenada gerada em 2 para incluir os pontos de alinhamento causal encontrados no final da etapa 5 ([34] cap. 2).

3.3 VQM_VFD

O VQM_VFD é um modelo de avaliação de qualidade vídeo que traduz devidamente o impacto perceptual do atraso variável de frames. Este modelo usa características perceptuais extraídas de blocos espaciais e temporais (S-T). Isto permite ao VQM_VFD sequenciar a qualidade subjectiva sobre uma grande variedade de distância de visualização e tamanho de imagem. O VQM_VFD obtém uma correlação de 0.9 para uma avaliação de qualidade subjectiva com tamanhos de imagem QCIF a HDTV.

3.3.1 Parâmetros VFD

Na secção 3.2 foi descrita uma técnica *Full Reference* para estimar o atraso variável de frames (VFD). VFD é uma técnica que usa a já apresentada na secção 3.2 mas vai mais longe através da exploração dos parâmetros de qualidade baseados em percepção que podem ser extraídos da informação VFD.

Existem dois parâmetros de qualidade objectivos que podem ser derivados da informação VFD. Um dos parâmetros usa a informação VFD exclusivamente enquanto que o outro usa a informação VFD em conjunção com a informação de movimento da cena. Ambos os parâmetros

foram empiricamente desenvolvidos com o objectivo de maximizar a sua correlação com as avaliações subjectivas dos testes de vídeos que possuem entre 8 a 15 segundos.

A técnica aqui descrita utiliza o algoritmo VFD para determinar a melhor correspondência entre um frame de vídeo original e processado. Em particular, esta técnica utiliza um array a que chama *Final_Fuzzy_Causal_Index* gerado pelo algoritmo VFD. O *Final_Fuzzy_Causal_Index* é uma matriz 2D onde o número de colunas é igual ao número de frames no vídeo processado, e o número de linhas é igual ao número dos frames originais que foram analisados para tentar fazer a correspondência a cada frame processado. Cada coluna da matriz 2D oferece um conjunto ordenado de alinhamentos de frames originais, do mais provável ao menos provável. Desta forma, o primeiro elemento de cada vector da coluna dá-nos o melhor índice do frame original para o frame processado. Os resultados para cada frame processado apenas incluem possíveis correspondências (índices do frame original que mais se assemelham ao frame processado). O array *Fuzzy* escreve-se da seguinte maneira:

$$Fuzzy(o_p, p) : p = 1, 2, \dots, N; o_p \in \{o_p^1, o_p^2, \dots, o_p^{L_p}\} \quad (3.1)$$

em que $p = 1, 2, \dots, N$ (o número total de frames no vídeo processado) e cada frame processado p corresponde a um conjunto o_p de L_p frames originais, onde L_p é o número de alinhamentos possíveis para cada frame processado p . Quando a correspondência não é ambígua, $L_p = 1$. Quando a correspondência é ambígua, $L_p > 1$. Frames de vídeo em cenas processadas que contêm pouco movimento, grandes distorções, e/ou contribuições de frames originais múltiplos vão criar ambiguidades na correspondência temporal ([6] cap. 2).

3.3.1.1 Parâmetro 1 (Par1)

O primeiro parâmetro (Par1) é um parâmetro de qualidade objectivo que é extraído apenas da informação VFD. A ideia é penalizar saltos de frames anormais nos índices de correspondências de frames originais o_p VFD da equação 3.1. Um salto de frame constituiria um salto de mais de um frame visto que os frames processados p avançam de 1 a N . A incerteza dos alinhamentos temporais pode ser acomodada calculando o salto de frames mínimo requerido que satisfaz o alinhamento temporal *Fuzzy* na equação 3.1. Para um dado frame p , primeiro calcula-se $Fuzzy_Max_Early(p)$, o índice de frames originais máximo no momento p , da seguinte forma:

$$Fuzzy_Max_Early(p) = \min(\max\{o_p^1, o_p^2, \dots, o_p^{L_p}\}, o_{p+1}^1) : p = 1, 2, \dots, N - 1 \quad (3.2)$$

O mínimo na equação 3.2 é usado para limitar o índice de frames originais máximo no momento p para ser menor que ou igual ao mais provável índice do frame no instante $p + 1$ (para causalidade). Ou seja, a incerteza nos alinhamentos de frames originais para o frame processado p não é permitido exceder o alinhamento mais provável para o frame $p + 1$. De seguida, para um determinado frame processado $p + 1$, calcula-se $Fuzzy_Min_Late(p + 1)$, o mínimo índice de frame original no momento $p + 1$, da seguinte forma:

$$Fuzzy_Min_Late(p + 1) = \max(\min\{o_{p+1}^1, o_{p+1}^2, \dots, o_{p+1}^{L_{p+1}}\}, Fuzzy_Max_Early(p)) \quad (3.3)$$

O máximo na equação 3.3 é usado para limitar o índice de frames originais no momento $p + 1$ para ser maior ou igual ao $Fuzzy_Max_Early(p)$. De seguida pode-se calcular os Saltos Anormais de Frames (AFJ - *Abnormal Frame Jumps*) no momento $p + 1$ como uma diferença entre as equações 3.3 e 3.2, em que $p = 1, 2, \dots, N - 1$:

$$AFJ(p + 1) = \max(0, [Fuzzy_Min_Late(p + 1) - Fuzzy_Max_early(p) - 1]) \quad (3.4)$$

Nesta equação, antes de se maximizar com 0, é subtraído 1 à diferença, uma vez que indica uma progressão normal de frames ao longo do tempo. É de notar a equação 3.4 não penaliza pausas de frames. Em vez disso, penalizações ocorrem quando o frame processado não é actualizado para um novo frame, o que é percebido como um salto anormal no movimento de uma cena. Pausas de frames mais longas produzem valores mais altos de AFJ quando o frame processado é finalmente actualizado.

O Parâmetro de qualidade vídeo $Par1$ é então calculado como:

$$Par1 = \log_{10} \left(1 + \sqrt{\text{mean}(AFJ^2)} \right) \quad (3.5)$$

AFJ^2 representa o quadrado elemento-a-elemento da série de tempo expressa pela equação 3.4.

É adicionado 1 à raiz quadrada da média do quadrado para evitar o logaritmo de zero, que também tem o objectivo de limitar a fronteira inferior do parâmetro a zero. Saltos de frame grandes que resultam em longos períodos de pausa de um frame são penalizados mais fortemente que muitos saltos anormais de frames pequenos uma vez que o quadrado de valores altos tem um efeito proporcional na média geral ([6] cap. 2.1).

3.3.1.2 Parâmetro 2 (Par2)

O Parâmetro descrito na secção anterior usa apenas informação VFD. Desta forma, os frames que foram descartados ou pausados irão ter o mesmo resultado, ou distorção, qualquer que seja a quantidade de movimento na cena. Por exemplo, uma transmissão a 10 frames por segundo (fps) de uma cena original de 30 fps (1 frame de cada 3 é transmitido) vai penalizar cenas com pouco e muito movimento igualmente. Um melhoramento a *Par1* incorporaria um balanceamento baseado na quantidade de movimento para que dessa forma mais movimento produziria parâmetros com maior valor do que cenas com pouco movimento. No limite, cenas sem movimento ou praticamente sem movimento apenas deveriam ser penalizadas ligeiramente quando os frames são descartados uma vez que estes frames são pouco notados. Uma possível função de informação temporal (TI - *temporal information*) no momento $p + 1$ é calculada como:

$$TI(p + 1) = \log_{10} \left(1 + \sqrt{\text{mean}_{over i,j} \{ [Y_{p+1}(i, j) - Y_p(i, j)]^2 \}} \right), (i, j) \in SROI \quad (3.6)$$

em que $p = 1, 2, \dots, N - 1$. Em 3.6 os cálculos são realizados usando pixels activos (i, j) , ou pixels que contêm uma real informação de vídeo, dentro da região espacial de interesse (SROI - *Spatial Region of Interest*). As vantagens da formulação 3.6 incluem uma função de peso zero para frames sem movimento (uma vez que o logaritmo de 1 é 0) e compressão do largo alcance dinâmico que resulta da operação da raiz quadrada da média do quadrado.

O parâmetro de vídeo *Par2* que usa informação VFD e informação de movimento pode ser calculada como:

$$Par2 = \log_{10} \left(1 + \sqrt{\text{mean}([AFJ \times TI]^2)} \right) \quad (3.7)$$

$[AFJ \times TI]^2$ representa uma multiplicação e quadrado elemento a elemento das duas séries temporais dados pelas equações 3.4 e 3.6. Tal como em 3.5, é adicionado 1 à raiz quadrada da

média para evitar o logaritmo de zero, que também limita a fronteira inferior do parâmetro a zero ([6] cap. 2.2).

3.3.1.3 Correlação dos parâmetros VFD

Esta secção examina a correlação entre os parâmetros *Par1* e *Par2* e qualidade subjectiva. O objectivo é avaliar em que medida a variância do MOS pode ser explicada por *Par1* e *Par2*. Três conjuntos de dados foram examinados, cada um contendo sequências de vídeo enviadas através da Internet, enviados a uma destas resoluções: QCIF (176 x 144), CIF (352 x 288) e VGA (640 x 480). Estas sequências de vídeo continham atrasos de frame variáveis.

O número de vídeos QCIF, CIF e VGA era 1964, 2143 e 1766 respectivamente. Cada conjunto foi submetido entre 14 a 17 testes subjectivos. O MOS de cada teste subjectivo era mapeado para a escala [0,1]. Para a escala, 0 é considerada o melhor resultado (melhor qualidade) e 1 o pior resultado (pior qualidade).

Os gráficos de dispersão que mostram os parâmetros *Par1* e *Par2* em relação à qualidade subjectiva na escala [0,1] são mostrados na Figura 3.2. Os coeficientes Pearson são dados no título dos gráficos. Como o quadrado do coeficiente de correlação Pearson é a percentagem da variância que é explicado pelos parâmetros, *Par1* explica 30.8%, 16.3% e 20.6% da variância subjectiva nos conjuntos QCIF, CIF e VGA respectivamente. *Par2* explica 38.1%, 20.2% e 23.9% da variância subjectiva nos conjuntos QCIF, CIF e VGA respectivamente.

Ambos os parâmetros têm uma resposta relativamente linear a mudanças em qualidade subjectiva. Os resultados de correlação de *Par2* são um pouco melhores que os de *Par1* em todos os conjuntos. Isto não implica necessariamente que *Par2* seria uma melhor adição a um conjunto de métricas objectivas que compõe um modelo que qualidade perceptual. Os dois parâmetros tendem a errar no sentido de detectar distorções por defeito em vez de detectar por excesso. Uma vez que estes parâmetros apenas medem distorções temporais, este comportamento é esperado ([6] cap. 3).

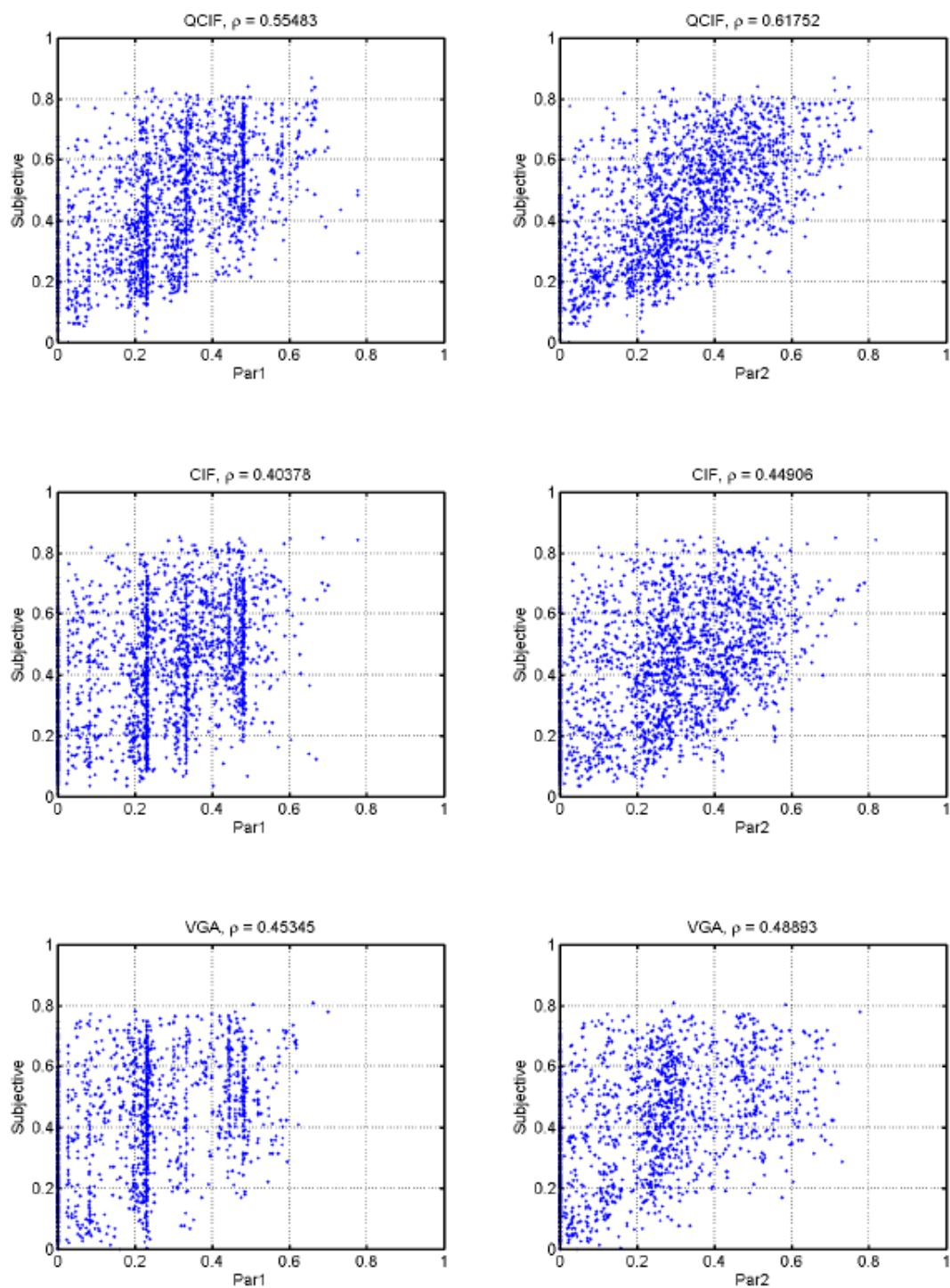


Figura 3.2: Qualidade subjectiva vs traçado objectivo para Par1 e Par2 ([6] Fig. 2)

3.3.2 Video Quality Model para Variable Frame Delay (VQM_VFD)

Na secção 3.2 descreve-se uma técnica FR para estimar a informação de atraso variável de frames (VFD). Em 3.3.1 vai-se mais longe propondo vários parâmetros baseados na percepção da qualidade que podem ser extraídos da informação VFD. Nesta secção é proposto o conjunto completo de parâmetros de qualidade VFD baseados na percepção.

Na Figura 3.3 é apresentada uma visão geral do sistema VQM_VFD. O vídeo processado é calibrado para remover as alterações de ganho e nível, de alterações espaciais e dimensionamento. De seguida a informação VFD (a melhor correspondência entre frame original e processado) é calculada e usada para mudar a ordem dos frames na sequência de vídeo original de forma a que corresponda aos frames do vídeo processado. Por exemplo, se o vídeo processado se repetisse frame sim frame não, a sequência original ajustaria-se a este comportamento. A informação VFD gerada neste passo, o vídeo processado calibrado e o vídeo original são depois usados para extracção de parâmetros baseados na percepção. Uma vez que todos os efeitos perceptuais de variações de atrasos de frames (pausas, saltos, etc) são removidos do vídeo original (que será usado como referência para FR VQM), os parâmetros de qualidade são também extraídos da própria informação VFD. Uma rede neural é usada para mapear todos os parâmetros VFD baseados na percepção para uma estimativa de qualidade ([7] cap. 1).

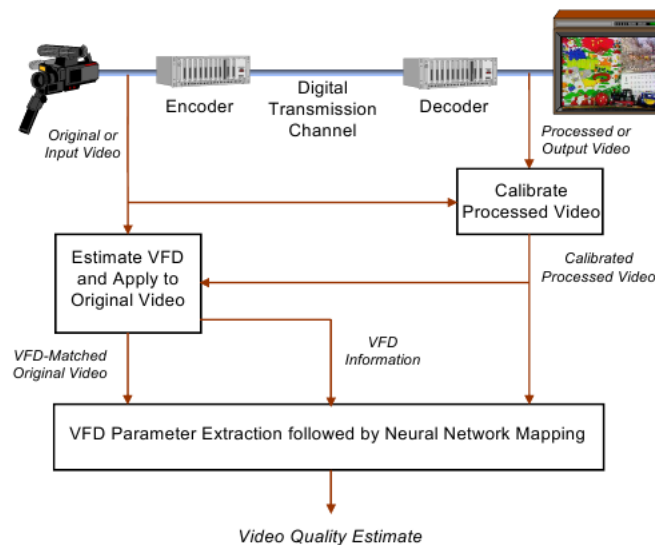


Figura 3.3: Esquema geral do VQM_VFD ([7] Fig. 1)

3.3.2.1 Parâmetros VFD baseados na percepção

Características e parâmetros extraídos de blocos espaciais e temporais (ST), como os mostrados na figura 3.4, foram usados para arquitectar VQMs com excelente correlação a qualidade subjectiva tal como o descrito na secção 2.6. Uma característica é uma quantidade de informação associada ou extraída de um bloco ST. Um parâmetro é uma medida de distorção de vídeo que resulta de comparar duas sequências de vídeo, a original e a processada.

Para vídeo com uma definição normalizada mostrado a uma distância de visualização (D) de 4 a 6 vezes a altura da imagem (H), estudos foram realizados para determinar o tamanho ideal dos blocos ST para efeitos de avaliação de qualidade. A altura H da imagem é a extensão vertical total da imagem. A distância ideal D simula o poder de visão do olho humano que não identifica pixels individuais.

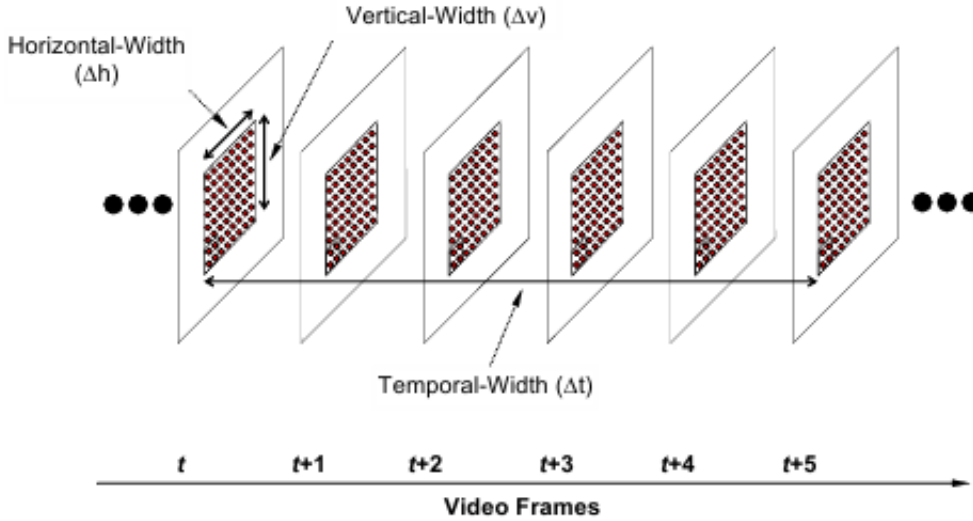


Figura 3.4: Pixels contidos num bloco ST ([7] Fig. 2)

Para produzir parâmetros de blocos ST baseados em física da visualização, é preciso ajustar o alcance angular θ subentendido no bloco ST pelo olho humano. Usando um alcance angular fixo θ para determinar o tamanho do bloco permite arquitectar um VQM que funciona sobre um grande variedade de tamanho de imagem e distancias de visualização. O alcance angular θ é escolhido para extrair todo o potencial das características e parâmetros extraídos (a sua correlação a qualidade subjectiva). Aparenta haver pouco valor em usar um alcance angular menor que 0.2 a 0.4 graus. Desta forma o VQM_VFD usa 0.4 graus.

Assumindo blocos quadrados e a aproximação do ângulo, o alcance vertical e horizontal Δ dos blocos ST (em pixels) é dado por:

$$\Delta h = \Delta v = N_V \times D \times \theta \times \frac{\pi}{180} \text{ pixels}, \quad (3.8)$$

onde N_V é o número total de pixels da imagem na posição vertical, D é a distância da imagem (em alturas de imagem H), e θ é o alcance angular dos blocos ST (em graus). Em relação à dimensão temporal do bloco ST (Δt), a visão humana requer uma sequência de frames de maneira a formar uma opinião. Desta forma, o VQM_VFD usa 0.2 segundos como dimensão temporal dos blocos ST.

O VQM_VFD usa 8 parâmetros de qualidade que são descritos de seguida:

- **HV_Loss** - O parâmetro *HV_Loss* do modelo VQM_VFD é derivado do General VQM descrito em detalhe no capítulo 6.3 em [33], mas com 4 diferenças:
 1. A correcção VFD é realizado na luminância do canal Y do vídeo original antes da extracção de características. Assim haverá um campo *vfd* adicionado ao nome do parâmetro Y.
 2. Filtros de detecção de margens horizontais (H) e verticais (V) únicos são usados em vez de filtros de tamanho fixo 13x13. Esta extracção de característica será chamada *hvA*
 3. O tamanho do bloco ST é dado por um alcance angular (0.4 graus) e uma duração de tempo (0.2 segundos) em vez de um número de pixels e frames definidos rigorosamente. Será denotado como o parâmetro *0.4deg_0.2s*.
 4. O tamanho de blocos ST são avaliados na forma de uma função quadrática pelo produto do nível do canal Y e a raiz da média quadrada (rms) do nível da informação temporal absoluta do bloco. Esta avaliação do bloco é realizada depois da função de comparação de blocos (*ratio_loss*) mas antes das funções de sondagem espacial e temporal (ex. *below5%_mean*).

O parâmetro HV_Loss do modelo VQM_VFD é dado por:

$$\begin{aligned} HV_Loss \\ = Y_vfd_hvA_0.4deg_0.2s_mean_3_ratio_loss_below5\%_mean_square_clip_0.06 \end{aligned} \quad (3.9)$$

Uma descrição das regras de nomenclatura dos parâmetros pode ser encontrada na secção 5.6 de [33].

- **HV_Gain** - O Parâmetro *HV_Gain* do modelo VQM_VFD é similar ao usado pelo modelo General VQM detalhado em [33]. O *HV_Gain* usa funções de sondagem de erros temporais e espaciais diferentes (diferenças 1 a 3 do parâmetro HV_Loss definidas anteriormente). O parâmetro *HV_Gain* é dado por:

$$HV_Gain = Y_vfd_hvA_0.4deg_0.2s_mean_3_log_gain_rms_rms \quad (3.10)$$

- **SI_Loss** - O Parâmetro *SI_Loss* do modelo VQM_VFD é semelhante ao usado no modelo geral VQM descrito em detalhe em [33]. O *SI_Loss* usa funções de sondagem de erros temporais e espaciais diferentes (diferenças 1 a 3 do parâmetro HV_Loss). O parâmetro *SI_Loss* é dado por:

$$SI_Loss = Y_vfd_siA_0.4deg_0.2s_std_12_ratio_loss_mean_above90\% \quad (3.11)$$

Em 3.11, o parâmetro *siA* denota o mesmo uso dos filtros de *hvA*, mas é usada a magnitude da informação espacial (SI).

- **SI_Gain** - O Parâmetro *SI_Gain* do modelo VQM_VFD é semelhante ao usado no modelo geral VQM detalhado em [33]. O *SI_Gain* usa funções de sondagem de erros temporais e espaciais diferentes (diferenças 1 a 3 do parâmetro HV_Loss). O parâmetro *SI_Loss* é dado por:

$$SI_Gain = Y_vfd_siA_0.4deg_0.2s_std_8_log_gain_above98\%tail_rms \quad (3.12)$$

- **TI_Gain** - O Parâmetro *TI_Gain* usa a mesma característica descrita no parâmetro *HV_Loss* para calcular a *rms* da energia de movimento ou informação temporal de um bloco ST. Este parâmetro calculado aqui não tem uma grande sensibilidade a frames descartados ou repetidos uma vez que estes são compensados pelo processo de correspondência VFD. Em vez disso, O *TI_Gain* mede distorções adicionadas no vídeo processado que não são compensadas pela correção VFD. O Parâmetro *TI_Gain* é dado por:

$$TI_Gain = Y_vfd_ti_0.4deg_0.2s_rms_3_log_gain_STabove95\%tail \quad (3.13)$$

- **RMSE_Gain** - O Parâmetro *RMSE_Gain* é um parâmetro *full reference* que é calculado como o erro *rms* entre blocos ST no vídeo processado e vídeo original dado pelo VFD. O *RMSE_Gain* é dado por:

$$RMSE_Gain = Y_vfd_fr_0.4deg_0.2s_rms_STmean \quad (3.14)$$

Em 3.14, *fr* denota um cálculo *full reference*, que é a diferença entre os pixels originais e os processados. O *RMSE_Gain* é a média rms de erro dos blocos ST individuais.

- **VFD_Par1** - Todos os parâmetros descritos anteriormente nesta secção usam uma sequência de vídeo original que foi correspondido com uma sequência de vídeo processado. Nenhum destes parâmetros pode quantificar o impacto perceptual de frames repetidos ou descartados. *VFD_Par1* captura o impacto perceptual dos frames repetidos ou descartados usando a informação VFD mostrada na Figura 3.3. *VFD_Par1* foi descrito na secção 3.3.1.1 está documentado detalhadamente no capítulo 2.1 de [6].
- **VFD_Par1.PSNR_VFD** - Quando VFD está presente na sequência de vídeo processada, o PSNR tem problemas que limitam o seu uso para avaliações subjectivas. O PSNR calculado depois da sequência de vídeo original ter sido emparelhada com a sequência processada ultrapassa estes problemas ([6] cap. 2.2). Um parâmetro que é calculado como o produto de *VFD_Par1* e *PSNR_VFD* captura os atributos perceptuais de ambos.

Os oito parâmetros de qualidade descritos devem ser mapeados para estimativas de qualidade subjectivas. Isto é feito através de uma rede neural. O algoritmo *Iterative Nested Least Squares* foi usado para mapear resultados subjectivos para a escala nominal (0,1). Isto permitiu que o conjunto de testes efectuados fosse usado para desenvolver e testar o mapeamento dos resultados.

3.3.2.2 Correlação do VQM_VFD

Existem diferentes maneiras de medir o desempenho de um modelo VQM. Apresentar-se-ão os gráficos de dispersão de resultados subjectivos face aos resultados objectivos e coeficiente

de correlação Pearson. A interpretação do coeficiente de correlação Pearson (p) é que o seu quadrado (p^2) oferece uma estimativa da variância MOS que é explicada pelo modelo objectivo.

Os resultados do desempenho do modelo VQM_VFD são apresentados separadamente para cada conjunto de dados, QCIF, CIF, VGA, SD e HD na Figura 3.5. Os coeficientes de correlação Pearson são 0.91, 0.91, 0.90, 0.91 e 0.90 respectivamente. É digno de nota atingir um coeficiente maior ou igual a 0.9 com um modelo de qualidade de vídeo sobre tão variados tipos de vídeo. Além disso, a resposta do VQM_VFD é estável uma vez que existem poucos pontos distantes da norma.

Na Figura 3.5 os resultados subjectivos variam entre 0 (melhor qualidade) e 1 (pior qualidade). É de notar que o gráfico relativo a vídeo HD demonstra uma falta de amostras de pobre qualidade em relação aos outros conjuntos.

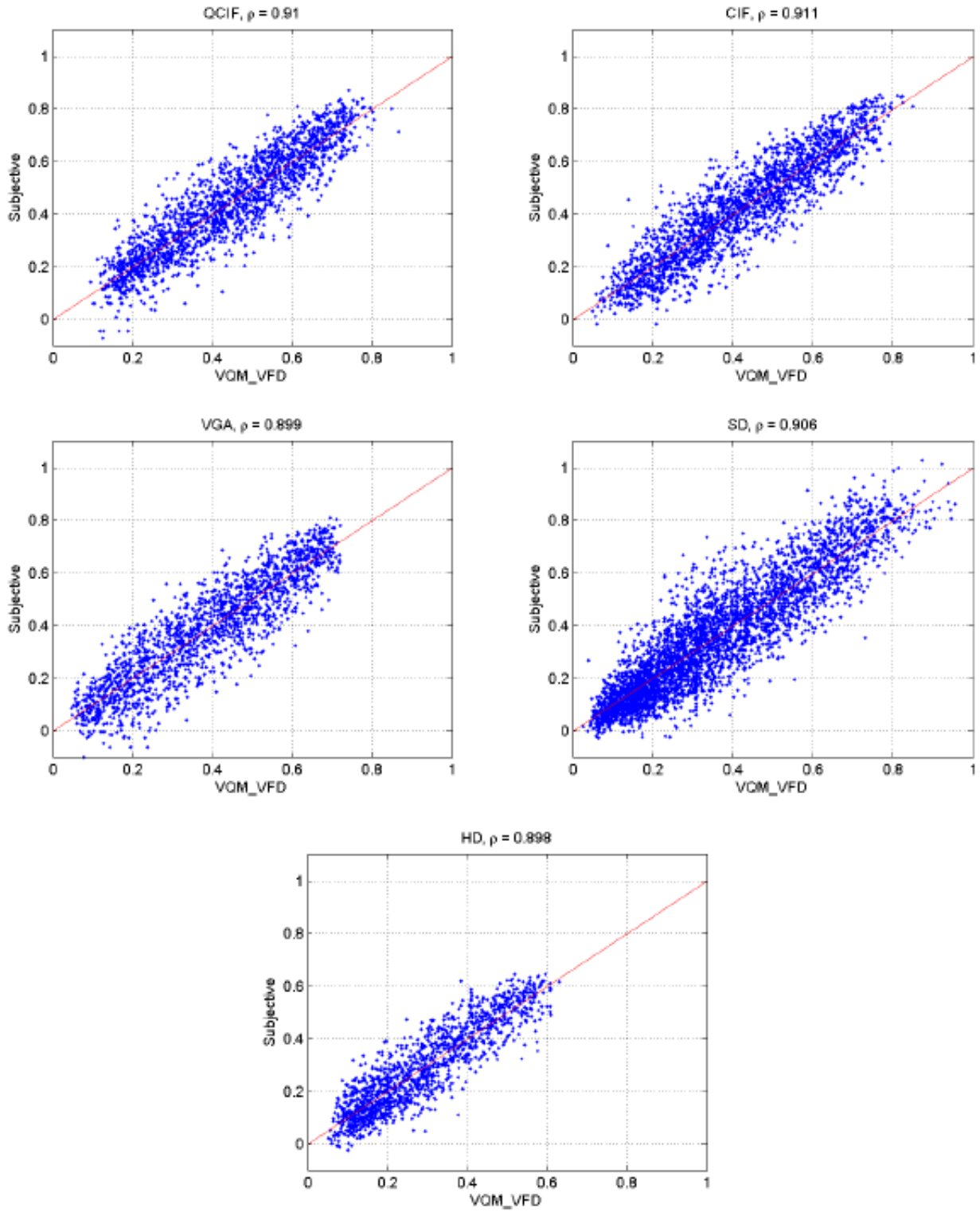


Figura 3.5: Desempenho do modelo VQM_VFD ([7] Fig. 5)

3.4 Resumo

Neste capítulo foi apresentado e detalhado a ferramenta e os métodos que irão ser usados para efectuar a avaliação vídeo. No próximo capítulo irá apresentar-se o sistema de testes montado e os procedimentos utilizados para efectuar avaliações de qualidade de experiência.

Capítulo 4

Arquitectura e Metodologia de testes

De forma a investigar o impacto que diferentes condições de operação de rede teriam na qualidade de experiência de um utilizador durante uma videochamada foi architectado um sistema de testes usando uma plataforma real. Este capítulo tem como objectivo oferecer uma visão detalhada sobre como foram efectuadas estas experiências.

4.1 Arquitectura do Sistema

A arquitectura do sistema consiste em dois computadores ligados em rede de forma a efectuar videochamadas entre os dois usando a aplicação SkypeTM. Um dos computadores possui duas placas de rede o que possibilita manter uma ligação à Internet partilhada com o segundo computador. O computador com duas placas de rede assume o papel de emissor da videochamada enquanto que o segundo computador assume o papel de receptor. A ligação entre emissor e receptor é controlada por um emulador de rede instalado no emissor, criando-se assim um ambiente controlado. A arquitectura do sistema é mostrada na Figura 4.1. Note-se que apesar de o sistema ter sido montado apenas com dois computadores, a independência dos vários componentes foi garantida, isto é, o emulador de rede que controla a ligação entre emissor e receptor é um elemento autónomo que apenas controla o tráfego direccionado do emissor para o receptor. O computador que funciona como receptor mantém uma ligação à Internet através do computador emissor não limitada pelo emulador de rede.

A colecta de dados para posterior análise pode ser dividida em quatro tarefas gerais:

1. configuração da rede com os parâmetros sujeitos a teste;
2. realização de uma videochamada através da rede configurada;
3. captura de áudio e vídeo provenientes do emissor;
4. análise das amostras de áudio e vídeo.

Os processos acima mencionados serão detalhados nas secções que se seguem.

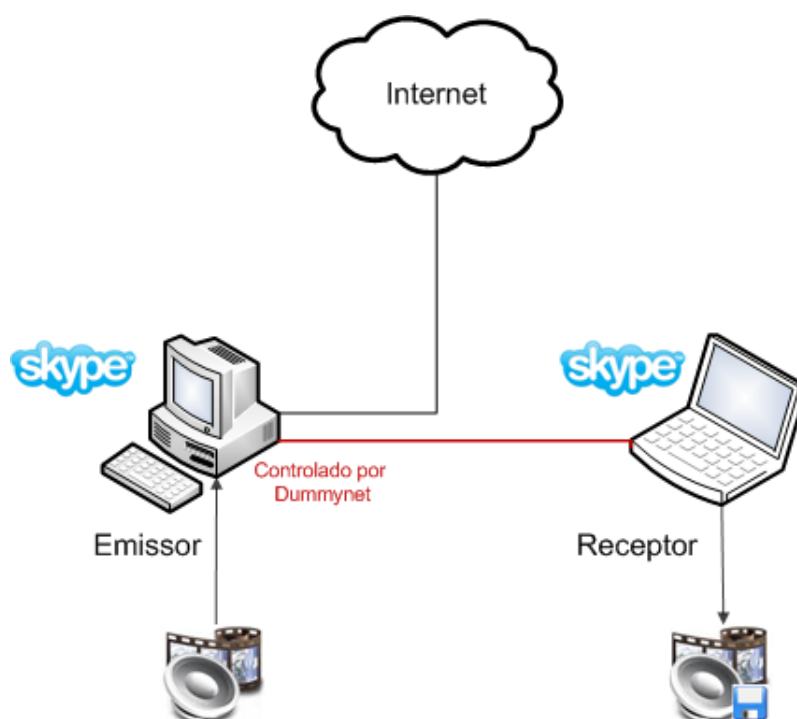


Figura 4.1: Arquitectura do sistema de testes

4.2 Emissão e Recepção de Vídeo/Áudio

A emissão e recepção de vídeos foi feita através da aplicação SkypeTM. De forma a criar um ambiente controlado na rede, isto é, uma ligação entre os dois computadores em que vários parâmetros da ligação são pré-definidos, foi usado o emulador Dummynet [35]. O Dummynet é uma ferramenta originalmente criada para testar protocolos de rede que permite introduzir parâmetros ajustáveis na rede tais como atraso, limitação de largura de banda e taxa de perda de

pacotes. O Dummynet pode ser instalado como um componente da placa de rede de um computador com o sistema operativo Windows e actua sobre o protocolo TCP/IP desse mesmo interface de rede, características que se encaixavam no sistema de testes previamente idealizado. Para efectuar testes de forma a simular uma carga de tráfego real decidiu-se optar pela utilização de um gerador de tráfego. Escolheu-se a ferramenta de geração de tráfego D-ITG (Distributed Internet Traffic Generator) [36] uma vez que permite gerar tráfego TCP e UDP variando o tamanho do *payload* dos pacotes e o tempo de partida entre pacotes, características que preenchiam os requisitos planeados.

Para a emissão de vídeo através do SkypeTM usou-se o software SplitCam [37]. O Splitcam permite a reprodução de um ficheiro de vídeo como a entrada de vídeo de uma videochamada. No receptor de vídeo utilizou-se o software SuperTintin [38] para a gravação do *stream* de vídeo proveniente do Skype.

Utilizou-se a ferramenta VirtualDub [39] para processamento de vídeo e áudio. O VirtualDub foi utilizado para tarefas tais como descompressão de vídeo, extracção da faixa áudio de um ficheiro de vídeo (a análise de áudio e vídeo só pode ser feita separadamente) e conversões entre formatos de áudio e vídeo.

4.3 Sequências de Vídeo e Áudio usadas

Todas as experiências planeadas foram executadas usando duas sequências de vídeo (ambas possuem áudio) distintas. Decidiu-se usar duas amostras de vídeo de forma a observar e comparar os resultados obtidos em duas cenas de vídeo com características distintas.

O primeiro vídeo usado (designado como "chart") possui um cenário típico de uma videochamada em que uma pessoa fala para uma câmara sem grande movimento em primeiro plano e um segundo plano estático. O vídeo encontra-se disponível em [40]. Para o segundo vídeo usou-se o vídeo denominado como "train" que se encontra em [41] (vídeo designado como "mobile" na página web em [41]). Este vídeo contém movimento em primeiro e segundo plano e uma grande variação de cores. A este vídeo foi adicionada (sobrepota com o vídeo) uma faixa de áudio usada para testar sistemas VoIP [42]. O primeiro vídeo tem uma duração de 10 segundos enquanto que o segundo tem uma duração de 8 segundos, sendo vídeos geralmente usados para testes com qualidade de experiência como se pode verificar pesquisando [40][41].

Ambos os vídeos foram convertidos para um *frame rate* de 25fps e uma resolução QVGA

(320x240) de forma a obter compatibilidade com o formato de gravação que o software Super-Tintin permite obter. O vídeo "chart" apresenta uma faixa de áudio com o formato PCM de 2 canais (*stereo*) e frequência de amostragem de 16kHz. O vídeo "train" tem uma faixa áudio de formato PCM de 1 canal (*mono*) e frequência de amostragem de 8kHz.

Apresenta-se na Figura 4.2 uma captura de ecrã do primeiro vídeo, e na Figura 4.3 apresenta-se uma captura de ecrã do segundo vídeo.



Figura 4.2: Captura de ecrã do vídeo "chart"



Figura 4.3: Captura de ecrã do vídeo "train"

4.4 Análise dos dados

Para analisar o conteúdo vídeo de forma a obter uma avaliação de qualidade, usou-se o programa BVQM [43], descrito na secção 3.1. A calibração usada em todos os vídeos foi a calibração *Full Reference* uma vez que proporciona maior precisão do que um algoritmo *Reduced Reference* e, de acordo com a situação de testes implementada, tinha-se acesso ao vídeo original e processado. O modelo escolhido para a avaliação de qualidade de vídeo foi o VQM_VFD, descrito na secção 3.3.2. O VQM_VFD foi lançado em finais de 2011 e é o modelo VQM mais preciso até à data.

O processo de análise de vídeo consistiu em:

1. descomprimir os vídeos gravados da situação de teste;
2. renomear os vídeos com a nomenclatura requerida pelo BVQM (descrito na secção 3.3);
3. seleccionar o vídeo original e os vídeos da situação de teste no BVQM e iniciar uma análise usando o modelo VQM_VFD com calibração *Full Reference*.

A análise das faixas áudio obtidas do vídeo foram analisadas usando uma implementação do algoritmo PESQ que faz parte da recomendação ITU-T [4], descrito na secção 3.3.2. As opções usadas na análise foram *narrowband listening* e a frequência de amostragem foi seleccionada conforme o valor presente nos ficheiros a analisar.

O processo de análise de áudio consistiu em:

1. extrair as faixas áudio dos vídeos gravados da situação de teste;
2. realizar a comparação entre a faixa de áudio do vídeo original e as faixas de áudio dos vídeos processados(o programa apenas permite analisar uma faixa de cada vez) usando o programa PESQ.

4.5 Resumo

Neste capítulo expôs-se a arquitectura do sistema de testes implementado bem como os componentes e procedimentos usados de forma a proceder-se à recolha e análise dos dados. No Capítulo 5 proceder-se-á a referir e explicar em detalhe as várias situações de teste projectadas e realizadas, apresentando-se em seguida os resultados obtidos.

Capítulo 5

Avaliação de QoE/QoS no Serviço de Videochamada

Neste capítulo detalham-se as várias situações de teste que foram executadas durante as quais se realizaram as videochamadas dando origem às gravações do conteúdo multimédia das mesmas. Os resultados das avaliações de qualidade das gravações foram recolhidos e expressos em gráficos. Neste capítulo apresentar-se-á também as respectivas avaliações e análises de cada situação de teste.

Como explicado no Capítulo 4 a recolha de dados que vão ser analisados é feita através da gravação do conteúdo recebido de uma videochamada. Na primeira série de testes foi implementada uma degradação artificial alterando certos parâmetros na rede que liga o computador emissor e receptor mostrados na Figura 4.1 usando o *Dummynet*. Tendo em conta três métricas diferentes, *delay* [0,50,100,150,200ms], *loss ratio* [0,5,10,15,20%] e *bandwidth* [ilimitada,100,75,50,25kbit/s] (atraso, taxa de perdas e largura de banda) emparelharam-se duas métricas diferentes em cada situação de teste enquanto que a métrica restante era definida como ideal [0ms,0%,ilimitada] i.e. sem qualquer atraso, sem perdas de pacotes e sem restrição da largura de banda. Os intervalos de valores foram escolhidos de forma a observar-se uma progressão razoável nos resultados obtidos, enquanto que o limite superior das métricas foi definido tendo em conta quando era atingido o valor máximo em termos de pior qualidade de experiência. A situação de largura de banda ilimitada corresponde à não imposição de um limite ao débito, sendo este apenas limitado pela capacidade da ligação (100Mbit/s). Esta tipologia de testes tem como objectivo testar o diferente impacto que estes parâmetros produzem na qualidade de experiência de uma videochamada.

A situação seguinte consistiu em introduzir *jitter* na rede usando o *Dummysnet*. Foram testados diferentes valores de *jitter*. Com este teste pretendia-se observar o efeito provocado pelo *jitter* na qualidade de experiência. Uma vez que se iria testar o efeito que um atraso constante nos pacotes provocaria decidiu-se então testar o efeito que um atraso variável entre pacotes impunha na qualidade de experiência. Os valores de *jitter* usados na situação de testes foram obtidos através de experiências preliminares tendo sido utilizado como valor máximo o valor de *jitter* com o qual se obteve uma qualidade de experiência próxima do pior limite de qualidade.

Na situação de testes final introduziu-se tráfego UDP e TCP separadamente na rede através do gerador de tráfego D-ITG. O propósito deste teste seria de avaliar os resultados de qualidade de experiência quando o sistema é submetido a uma situação aproximada de tráfego real intenso. A quantidade de tráfego máxima a que se decidiu submeter o sistema foi determinada através de testes iniciais tendo-se decidido usar como limite máximo um valor que se considerou uma carga bastante elevada em situações normais na qual o nível da qualidade já se aproximava do pior limite.

Ao todo foram gravadas mais de 390 amostras de vídeo e realizadas à volta de 780 avaliações de qualidade (vídeo + áudio). Como já foi referido anteriormente cada teste foi realizado com dois vídeos diferentes, o vídeo denominado por "chart" e o vídeo denominado "train". Estes dois vídeos possuem características diferentes, "chart" um vídeo com um fundo estático mais parecido com o que se pode esperar numa videochamada Skype ou teleconferência e "train" com um cenário movimentado e repleto de cor. Esperando-se com isto observar eventuais diferenças na qualidade de experiência usando tipos de vídeo diferentes. As análises de vídeo foram realizadas usando o BVQM com uma calibração *Full Reference* e usando o modelo VQM_vfd que tem como resultado uma pontuação QoE com um valor compreendido entre 0 e 1, sendo 0 o melhor resultado. As análises de áudio foram analisadas usando o PESQ que tem como resultado uma pontuação MOS compreendida entre 0 e 5.

De forma a oferecer uma melhor visualização sobre o conjunto de testes realizados, apresenta-se na Tabela 5.1 um resumo dos mesmos.

Conjunto de testes	Tipo de teste
1ª série	Impacto provocado por <i>delay/loss ratio/bandwidth</i>
2ª série	Efeito causado pelo <i>jitter</i>
3ª série	Testes com elevada carga de tráfego TCP/UDP

Tabela 5.1: Tipologia dos testes

5.1 Loss Ratio vs Delay

Na primeira experiência emparelharam-se as métricas *delay* e *loss ratio* enquanto que a largura de banda não foi limitada. Os valores usados na configuração do *delay* foram de 0, 50, 100, 150 e 200 milissegundos. Os valores usados na configuração do *loss ratio* foram de 0, 5, 10, 15 e 20 por cento. Os comandos usados no Dummynet foram os seguintes:

- `ipfw add pipe 1 ip from any to 192.168.0.1`
- `ipfw pipe 1 config plr [0-0.2] delay [0-200]ms`

Nos comandos apresentados anteriormente, "192.168.0.1" é o endereço IP do computador que recebe as videochamadas e "plr" é o *loss ratio*.

Para o vídeo "chart", de forma a obter uma boa referência e validação dos resultados obtidos, foram efectuadas 75 gravações de vídeo, 3 por cada parâmetro. Para o vídeo "train" para as mesmas métricas foram realizadas 25 gravações.

Na Figura 5.1 pode-se observar o gráfico que demonstra os resultados obtidos para a qualidade de vídeo nesta experiência para o vídeo "chart". Pode-se constatar que o *loss ratio* é um factor de maior importância na qualidade de vídeo que o *delay*, havendo até um momento no ponto LR 10% em que os vídeos com maior *delay* obtêm melhor pontuação. A partir de uma taxa de perdas superior a 10% a qualidade começa deteriorar-se significativamente. Na avaliação de áudio, mostrada na Figura 5.2 verifica-se o mesmo fenómeno notando-se no entanto uma pobre avaliação de qualidade mas uma queda de qualidade menos significativa.

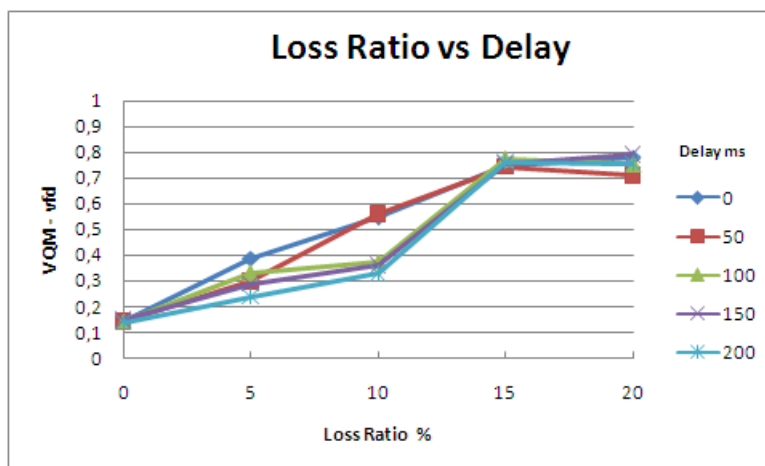


Figura 5.1: Qualidade vídeo "chart" - Loss Ratio vs Delay

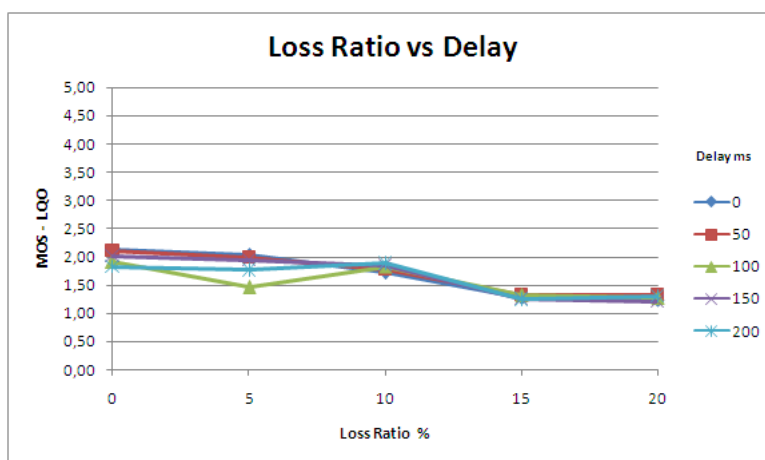


Figura 5.2: Qualidade áudio "chart" - Loss Ratio vs Delay

Os resultados obtidos na avaliação de vídeo usando o vídeo "train" revelaram-se idênticos aos obtidos no vídeo "chart" apesar de o vídeo "train" conter uma quantidade mais significativa de movimento. Nota-se no entanto uma maior flutuação entre valores neste vídeo devido a cada valor ser resultante de um menor número de avaliações que no caso do vídeo "chart". A qualidade áudio obteve no entanto uma ligeira melhor pontuação que o áudio do vídeo "chart". Os gráficos dos resultados vídeo e áudio para o vídeo "train" estão representados nas Figuras 5.3 e 5.4 respectivamente.

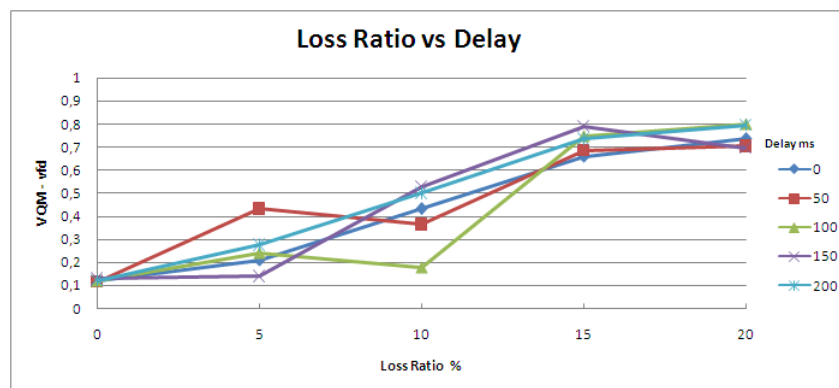


Figura 5.3: Qualidade vídeo "train" - Loss Ratio vs Delay

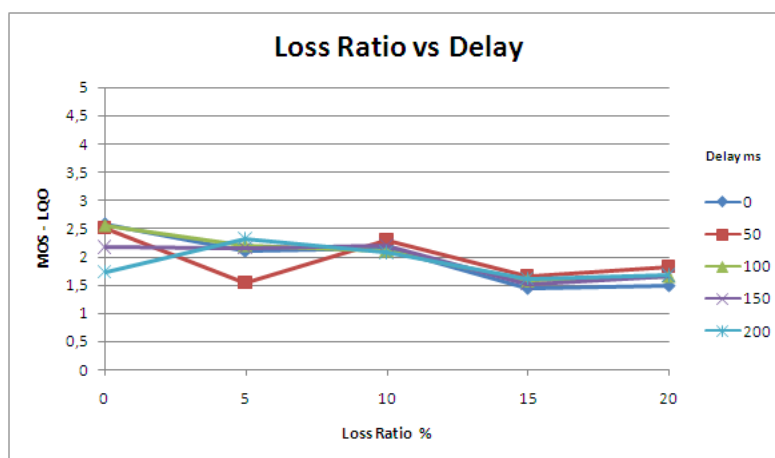


Figura 5.4: Qualidade áudio "train" - Loss Ratio vs Delay

5.2 Loss Ratio vs Bandwidth

Nesta experiência emparelharam-se as métricas de *loss ratio* e *bandwidth* enquanto que o atraso foi considerado nulo. Os valores usados na configuração do *loss ratio* foram de 0, 5, 10, 15 e 20 por cento. Os valores usados na configuração da largura de banda foram ilimitada, 100, 75, 50 e 25kbit/s. Os comandos usados no Dummynet foram os seguintes:

- `ipfw add pipe 1 ip from any to 192.168.0.1`
- `ipfw pipe 1 config plr [0-0.2] bw [0-100]Kbit/s`

Na Figura 5.5 pode-se observar o gráfico que demonstra os resultados obtidos para a qualidade de vídeo nesta experiência para o vídeo "chart". Pode-se constatar que ambas as métricas têm influência na qualidade de vídeo para os valores testados. Observa-se que uma largura de banda menor que 50kbit/s resulta numa qualidade diminuída instantaneamente. Pode-se verificar também que uma maior largura de banda ajuda na resolução de problemas de perdas de pacotes. No entanto a partir de um *loss ratio* maior que 10% volta a verificar-se uma degradação acentuada da qualidade de vídeo. Na avaliação de áudio, mostrada na Figura 5.6 obtém-se uma má avaliação para as condições de teste iniciais mantendo-se, no entanto, essa qualidade.

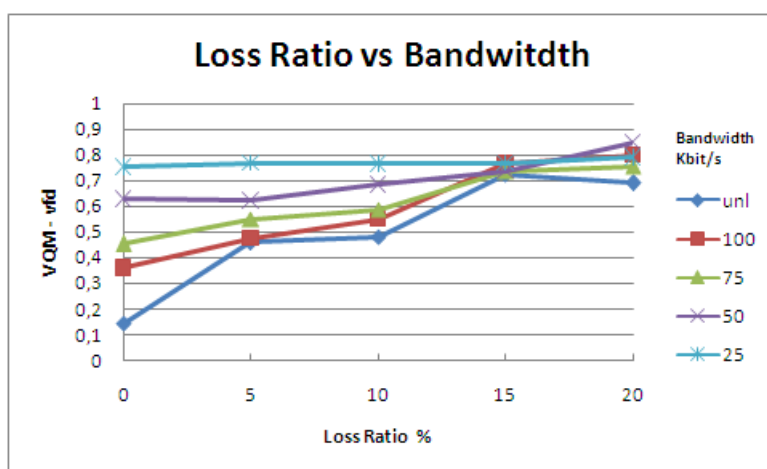


Figura 5.5: Qualidade vídeo "chart" - Loss Ratio vs Bandwidth

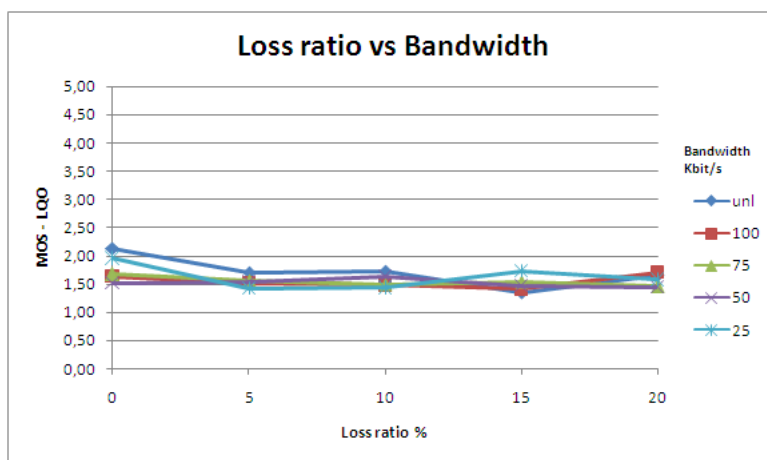


Figura 5.6: Qualidade áudio "chart" - Loss Ratio vs Bandwidth

Os resultados obtidos na avaliação de vídeo usando o vídeo "train" revelaram-se similares aos obtidos no vídeo "chart". No entanto é de notar que no vídeo "train" a qualidade já diminui de forma mais acentuada após uma taxa de perdas de pacotes acima dos 5%. A qualidade áudio obteve uma melhor pontuação que o áudio do vídeo "chart" o que estará relacionado com o facto de, à partida, o áudio do vídeo "train" ser de pior qualidade o que implica menos informação a transmitir. Os gráficos dos resultados vídeo e áudio estão representados nas Figuras 5.7 e 5.8 respectivamente.

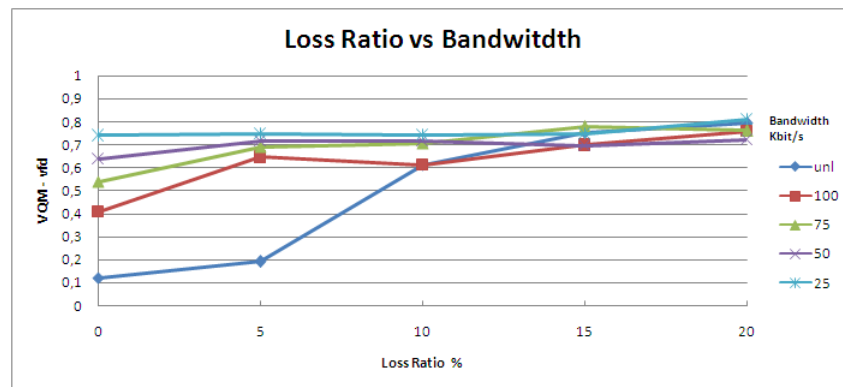


Figura 5.7: Qualidade vídeo "train" - Loss Ratio vs Bandwidth

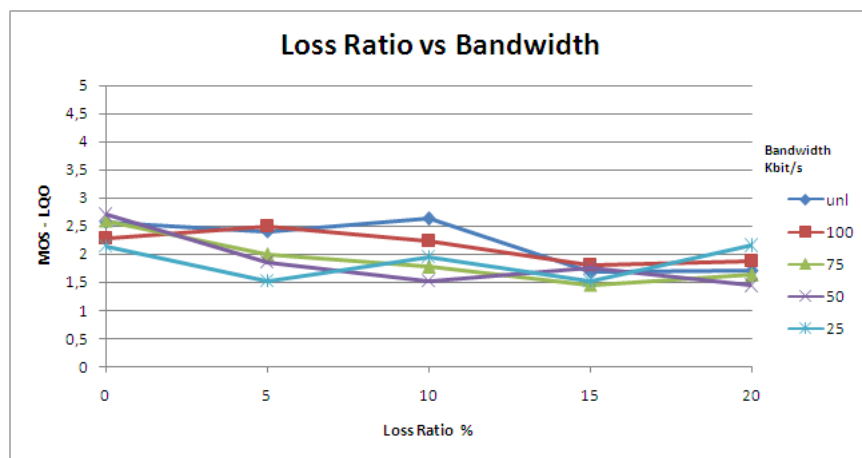


Figura 5.8: Qualidade áudio "train" - Loss Ratio vs Bandwidth

5.3 Bandwidth vs Delay

Nesta experiência estudam-se as métricas de *bandwidth* e *delay* para uma taxa de perdas nula. Os valores usados na configuração da largura de banda foram ilimitada, 100, 75, 50 e 25kbit/s. Os valores usados na configuração do atraso foram 0, 100, 200, 300 e 400ms. Decidiu-se alargar o limite superior do atraso a 400ms uma vez que nos testes iniciais com um limite de 200ms não se verificava influência negativa por parte do atraso. Os comandos usados no *Dummynet* foram os seguintes:

- `ipfw add pipe 1 ip from any to 192.168.0.1`
- `ipfw pipe 1 config delay [0-400]ms bw [0-100]Kbit/s`

Na Figura 5.9 apresenta-se o gráfico demonstrativo dos resultados obtidos para a qualidade de vídeo na experiência para o vídeo "chart". Pode-se constatar que o *delay*, apesar de chegar a valores como 400ms não exerce grande influência sobre a qualidade vídeo. A razão pela qual isto acontece deve-se provavelmente ao facto de o atraso ser constante para todos os pacotes de informação, isto é, uma vez que o atraso é constante, o vídeo no receptor sofrerá apenas de um atraso inicial relativamente à recepção da informação enviada não sofrendo no entanto em termos de qualidade uma vez que o conteúdo original permanece não degradado. Ao longo do gráfico a qualidade avaliada mantém-se constante, influenciada apenas pela largura de banda. Na avaliação de áudio, mostrada na Figura 5.10 observam-se algumas variações na pontuação de qualidade. A partir de um valor igual ou superior de atraso de 300ms, a avaliação áudio, parece emular os resultados já observados na qualidade vídeo com as linhas indicativas da pontuação VQM a indicarem alguma estabilidade.

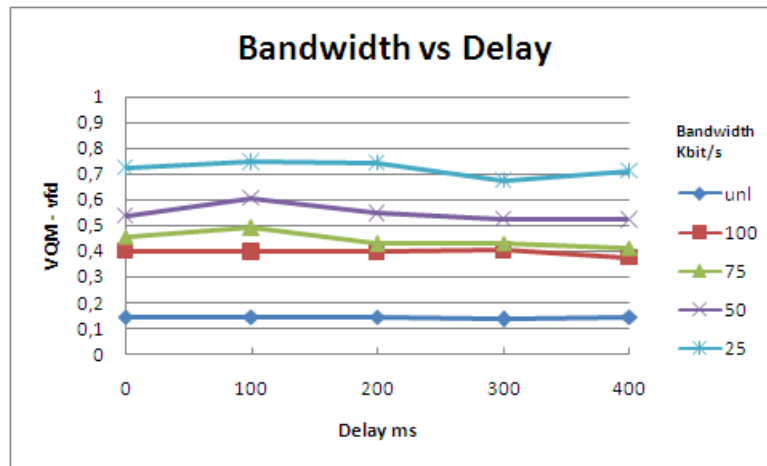


Figura 5.9: Qualidade vídeo "chart" - Bandwidth vs Delay

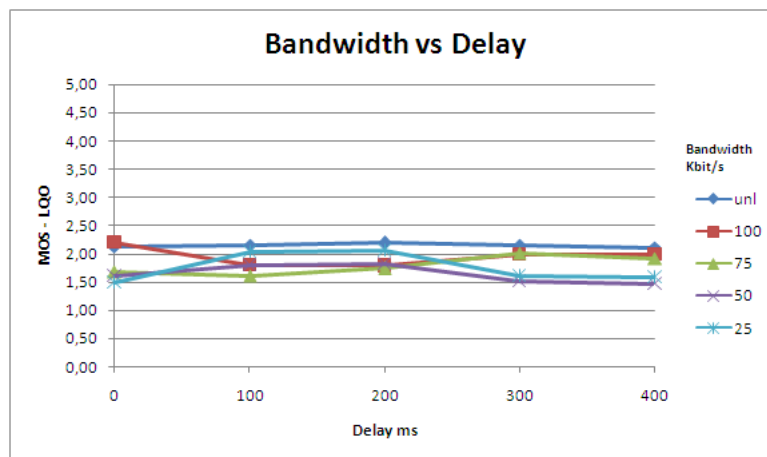


Figura 5.10: Qualidade áudio "chart" - Bandwidth vs Delay

Os resultados obtidos na avaliação de vídeo usando o vídeo "train" diferem substancialmente dos resultados obtidos com "chart". Na Figura 5.11 podemos verificar que os resultados da qualidade de vídeo são piores à partida que no vídeo "chart". A partir de um valor superior a 100ms a qualidade de vídeo em "train" deteriora-se significativamente contrariando a estabilidade pontual que acontecia no vídeo "chart". Esta deterioração de qualidade que não se verifica no vídeo "chart" poderá no entanto dever-se simplesmente à redução da largura de banda e não ao atraso, uma vez que o vídeo "train" é um vídeo que contém bastante movimento, logo necessitará de maior quantidade de informação para reconstruir uma imagem correctamente. A qualidade áudio no vídeo "train", apresentada na Figura 5.12, é também significativamente melhor que a avaliada no vídeo "chart". Podemos constatar que existe alguma flutuação das linhas indicativas

de qualidade e é de notar que terminam numa tendência ascendente. Isto poderá dever-se a algoritmos Skype que detectando uma má transmissão de vídeo provocada por condições adversas de rede opta por preservar a qualidade de áudio uma vez que o áudio tendo menor quantidade de informação é mais fácil de preservar.

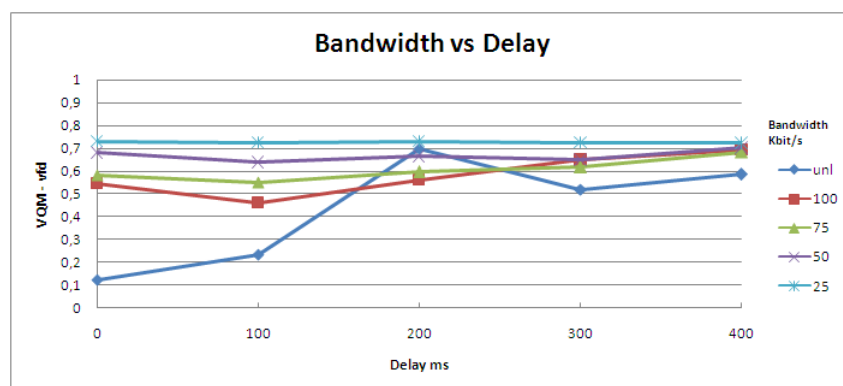


Figura 5.11: Qualidade vídeo "train" - Bandwidth vs Delay

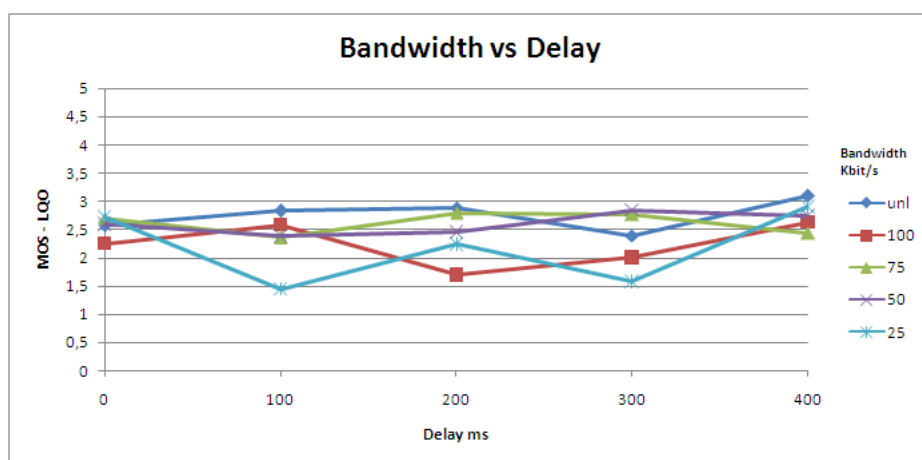


Figura 5.12: Qualidade áudio "train" - Bandwidth vs Delay

5.4 Jitter

Nesta experiência usou-se a aplicação *Dummysnet* para introduzir *jitter* na rede com o objetivo de verificar o efeito do *jitter* na qualidade de vídeo e áudio. A introdução de *jitter* através do *dummysnet* foi obtida da seguinte maneira: criaram-se vários *pipes* (cujo destino era o com-

putador receptor da chamada) com diferentes valores de atraso e de seguida adicionou-se uma probabilidade de escolha a cada *pipe*. Os comandos seguintes exemplificam este processo:

- `ipfw add prob 0.25 pipe 1 ip from any to 192.168.0.1`
- `ipfw add prob 0.25 pipe 2 ip from any to 192.168.0.1`
- `ipfw add prob 0.25 pipe 3 ip from any to 192.168.0.1`
- `ipfw add prob 0.25 pipe 4 ip from any to 192.168.0.1`
- `ipfw pipe 1 config delay 50ms`
- `ipfw pipe 2 config delay 100ms`
- `ipfw pipe 3 config delay 150ms`
- `ipfw pipe 4 config delay 200ms`

Os comandos acima resultam na criação de 4 *pipes*. Cada um destes pipes tem uma probabilidade de 25% de ser usado, ou seja, cada pacote destinado à máquina 192.168.0.1 tem uma probabilidade de 25% de seguir por um desses pipes atrasando-se o valor configurado nesse pipe. Desta forma consegue-se impor uma variação no atraso no tráfego gerado.

Em todos os testes o número de pipes foi 4, tendo cada um uma probabilidade de 25% de utilização. Os valores de atrasos configurados para cada pipe em cada situação de teste foram [50,100,150,200ms], [100,200,300,400ms] e [100,300,700,900ms]. Através da ferramenta Iperf [44] efectuaram-se 5 medições de *jitter* para cada conjunto de atrasos e efectuou-se a média de forma a obter-se um valor representativo (e aproximado) do *jitter* nesse caso. Com um conjunto [50,100,150,200ms] mediu-se um *jitter* médio de 74,69ms, em [100,200,300,400ms] mediu-se 127,93ms e para [100,300,700,900ms] um *jitter* médio de 300,94ms.

Na Figura 5.13 apresenta-se o gráfico com os resultados da experiência. Pode-se verificar que logo no salto de um sistema sem *jitter* para um *jitter* médio de aproximadamente 74ms, a qualidade diminui consideravelmente. A qualidade áudio, Figura 5.14, diminui mais lentamente que o vídeo tendo o maior decréscimo entre os dois maiores valores de *jitter*.

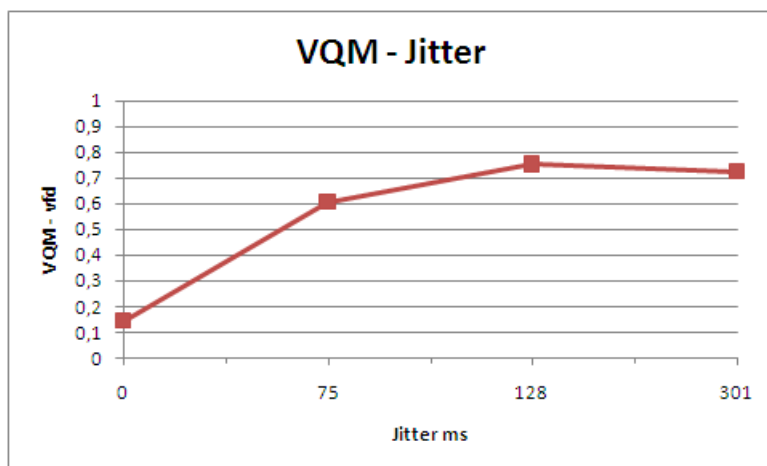


Figura 5.13: Qualidade vídeo "chart" - Jitter

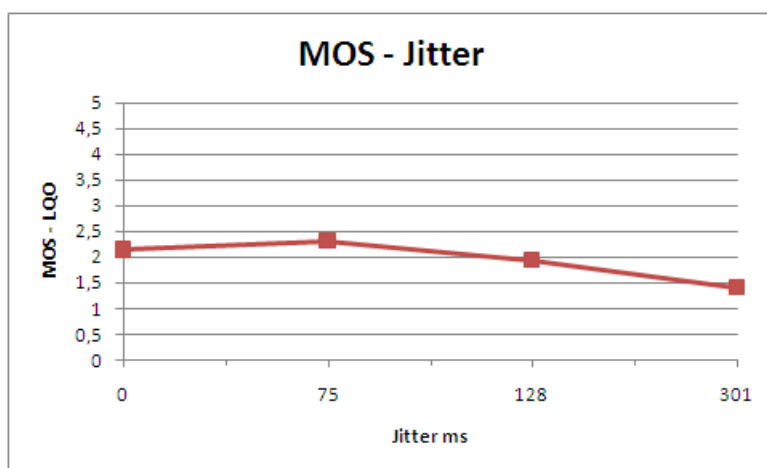


Figura 5.14: Qualidade áudio "chart" - Jitter

Usando o vídeo "train" nas mesmas condições originou um resultado em tudo similar ao do vídeo "chart" com um grande decréscimo de qualidade inicial e uma descida mais gradual nos pontos seguintes. Podemos observar este gráfico na Figura 5.15. O áudio "train", Figura 5.16, também se comportou de maneira similar ao do vídeo "chart", tendo um comportamento bastante bom no início e mantendo esse comportamento até chegar a um *jitter* médio de aproximadamente 300ms, altura em que decresce abruptamente.

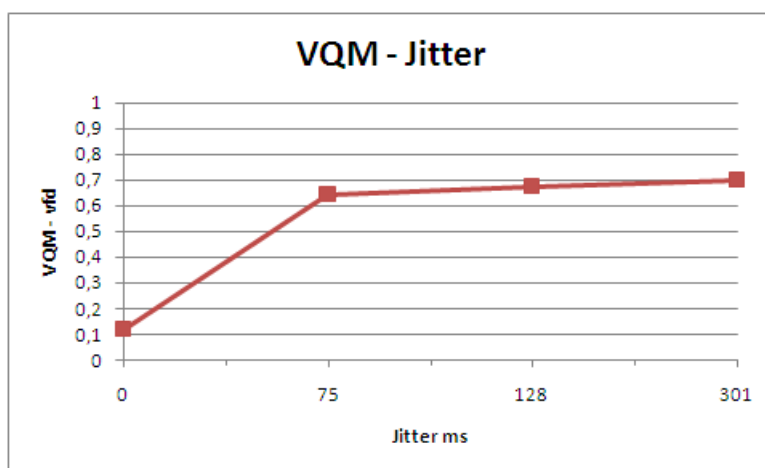


Figura 5.15: Qualidade vídeo "train" - Jitter

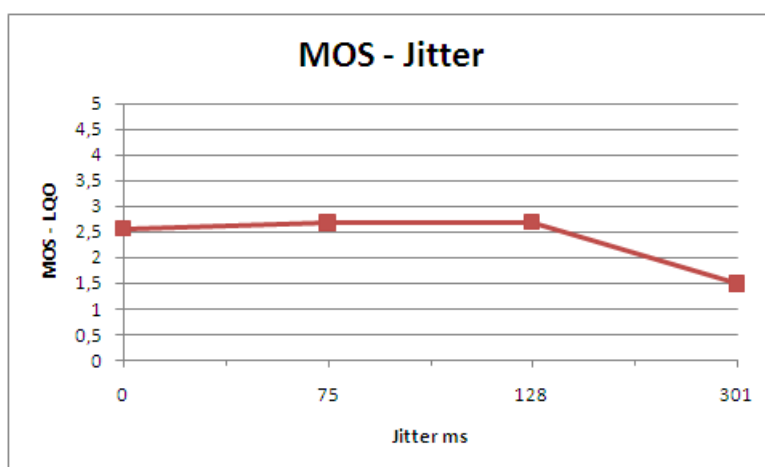


Figura 5.16: Qualidade áudio "train" - Jitter

5.5 TCP vs UDP

Com o objectivo de testar a qualidade vídeo quando submetida a uma situação de tráfego real intenso realizou-se uma série de testes utilizando um gerador de tráfego, nomeadamente o D-ITG. Usando a mesma arquitectura usada até agora, instalou-se a aplicação servidor do D-ITG no computador emissor de videochamadas e a aplicação cliente no receptor. A geração do tráfego fluía no sentido da aplicação servidor para o cliente. Com o objectivo de sobrecarregar a rede de forma a verificar-se o impacto na qualidade, decidiu-se limitar a largura de banda da ligação para

CAPÍTULO 5. AVALIAÇÃO DE QOE/QOS NO SERVIÇO DE VIDEOCHAMADA

15Mbit/s de modo a tornar esta tarefa exequível. Os comandos usados na geração de tráfego com o D-ITG foram:

- `-a 192.168.0.1 -rp 9003 -O 80 -o 1000048576 -T UDP -t 300000`
- `-a 192.168.0.1 -rp 9004 -O 80 -o 1000048576 -T UDP -t 300000`

Nos comandos acima exemplificados "-a" indica o endereço de destino, "-rp" a porta de destino, "-O" é o tempo de partida entre pacotes que seguirá uma distribuição de Poisson e toma como argumento a média de pacotes por segundo (convertidos para um intervalo de tempo de $\frac{1000}{x}$), "-o" o tamanho do *payload* em bytes que segue uma distribuição de Poisson, "-T" indica o protocolo a usar e "-t" a duração da geração de tráfego.

Foi decidido usar um *payload* grande (aproximadamente de 950MB) em cada pacote gerado para garantir que a transmissão de um pacote não acabaria durante a duração do teste e desta forma sobrecarregar sempre mais a rede. A decisão de ter um grande *payload* prende-se com o facto de se querer uma transmissão de informação contínua, isto é, saber sempre a quantidade de informação a ser transmitida. Com um *payload* reduzido, e por exemplo uma largura de banda não limitada, vários pacotes seriam eventualmente transmitidos na totalidade enquanto que isto não aconteceria na mesma medida numa situação de teste em que a largura de banda seria mais reduzida e desta forma a quantidade de informação a ser transmitida iria variar. Os únicos parâmetros que foram alterados nesta série de testes, foi o intervalo de tempo de partida entre pacotes e o tipo de protocolo que variava entre UDP e TCP. O intervalo de tempo de partida entre pacotes variou entre o conjunto de valores [0,80,300,1000] pacotes por segundo. Após de se realizarem vários testes com valores diferentes decidiu-se apresentar os resultados obtidos sobrepondo os resultados TCP e UDP para este conjunto de valores.

Como seria de esperar a qualidade de vídeo obteve melhores resultados quando o tráfego gerado era TCP, uma vez que este possui controlo de fluxo e de congestão. Na Figura 5.17 pode-se constatar isso mesmo, com o vídeo "chart" mesmo estando a rede com um nível de carga elevado a qualidade de vídeo parece estabilizar num valor de qualidade aceitável com o tráfego TCP enquanto que com UDP os resultados de qualidade são menos satisfatórios para o caso particular da videochamada. Todavia, a qualidade áudio para o vídeo chart, como se pode ver na Figura 5.18, mantém-se num nível aceitável e mantém-se estável durante toda a situação de teste tanto para TCP como para UDP.

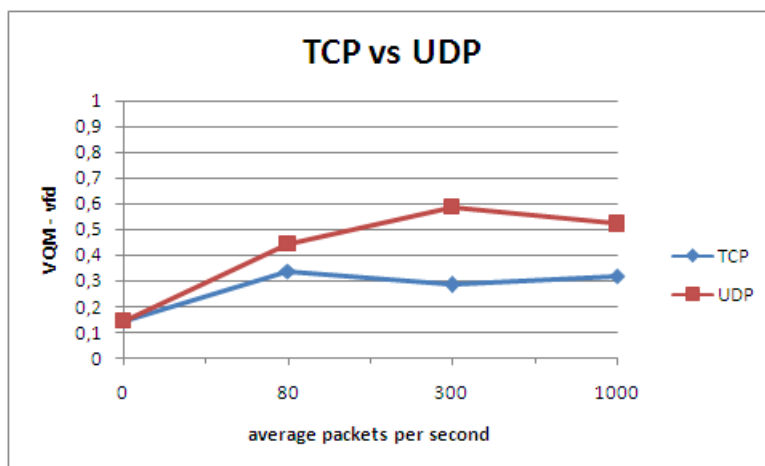


Figura 5.17: Qualidade vídeo "chart" - TCP vs UDP

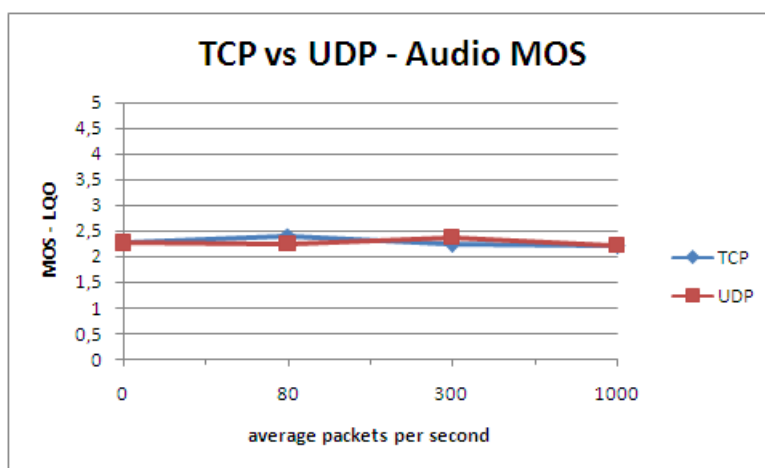


Figura 5.18: Qualidade áudio "chart" - TCP vs UDP

Com o vídeo "train" os resultados de qualidade de vídeo foram piores que no caso anterior verificando-se uma má pontuação desde o início da geração de tráfego. Com o tráfego TCP, a avaliação de qualidade manteve-se num bom nível, comparável ao caso do vídeo "chart". Os resultados encontram-se na Figura 5.19. Tal como no caso do vídeo "chart", os resultados áudio do vídeo "train" foram satisfatórios mantendo-se estáveis durante toda a duração dos testes para TCP e UDP, tendo obtido uma pontuação ligeiramente melhor que a do vídeo "chart". Os resultados áudio para o vídeo "train" estão disponíveis na Figura 5.20.

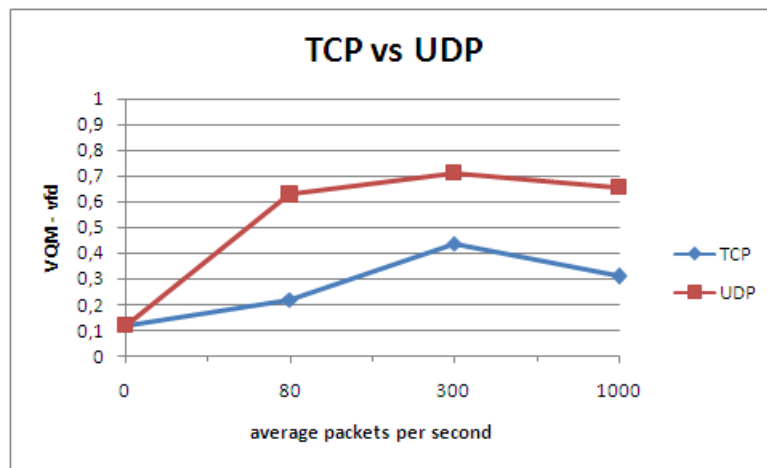


Figura 5.19: Qualidade vídeo "train" - TCP vs UDP

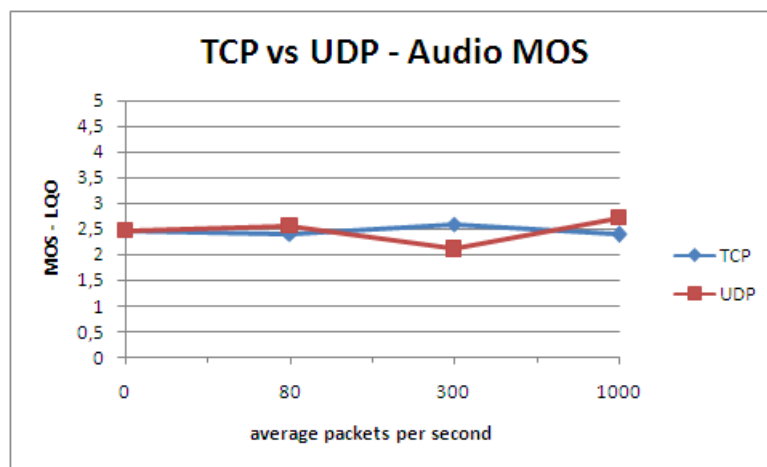


Figura 5.20: Qualidade áudio "train" - TCP vs UDP

5.6 Resumo

Neste capítulo foram descritos os vários testes efectuados no sistema de testes idealizado e apresentaram-se os resultados obtidos em cada situação e a interpretação dos mesmos. Prossegue-se no Capítulo 6 com as conclusões sobre os resultados do trabalho realizado.

Capítulo 6

Conclusões

Neste trabalho analisamos os efeitos que diferentes condições na rede produzem na qualidade de experiência durante uma videochamada (vídeo e áudio). Um sistema de testes foi arquitectado onde se realizaram várias séries de testes e onde se recolheram dados que permitiram efectuar a avaliação de qualidade de experiência usando a aplicação BVQM para avaliação de qualidade vídeo e o algoritmo PESQ para avaliação de qualidade áudio.

Na análise efectuada neste trabalho podemos verificar que existe uma relação directa entre QoS e QoE. Observando os resultados do conjunto de testes com os parâmetros atraso, largura de banda e taxa de perdas pode-se concluir que uma taxa de perdas superior a 10% tem um efeito de degradação bastante acentuado em qualquer situação. Uma largura de banda abaixo do 50kbit/s produzirá uma qualidade de vídeo pouco favorável. Dos 3 parâmetros o atraso parece ser o único que não exerce grande influência sobre a qualidade vídeo quando o vídeo tem um conteúdo estático (menos informação a transmitir) sendo que com atrasos bastantes largos, usando o vídeo "chart", (400ms) não produziu degradação visível (ao contrário dos testes com o vídeo "train") (secção 5.3). Ainda nesta série de testes conclui-se que a qualidade áudio apresenta maior imunidade às condições da rede que o vídeo, uma vez que é bastante menos informação a ser transmitida. Esperava-se no entanto melhores resultados nas avaliações áudio, não se obteve uma avaliação superior a razoável em nenhuma situação de teste. A aplicação Skype também parece favorecer a transmissão áudio em casos de pobres condições na rede como parece indicar também o facto de requerer ao utilizador terminar a transmissão vídeo mantendo apenas o áudio. É de notar também que no caso do vídeo "train" os resultados foram consistentemente melhores o que se deverá ao facto de o áudio do vídeo "train" ser de menor qualidade original que o áudio do vídeo "chart" (menos informação a ser transmitida).

Através da série de experiências com o *jitter* retira-se a conclusão que este é um factor com grande influência na qualidade de vídeo. Como se pode verificar na secção 5.4 um valor de *jitter* de 74ms já é suficiente para provocar um salto negativo considerável na qualidade de experiência, tornando-se o vídeo bastante degradado com um valor à volta dos 127ms. A qualidade áudio sofre menos com as variações de atraso mantendo-se num nível estável até se atingir um valor de *jitter* de 300ms altura em que piora significativamente a sua qualidade.

Na situação de testes de carga usando o gerador de tráfego para produzir um volume de tráfego TCP e UDP adequado, conclui-se que o tráfego TCP afecta menos a qualidade de experiência que o tráfego UDP, como seria de esperar tendo em conta os mecanismos de controle do TCP. A qualidade de vídeo manteve-se relativamente aceitável mesmo quando a rede estava sujeita a tráfego intenso. Observou-se no entanto uma deterioração gradual no vídeo nos testes com tráfego UDP. A qualidade áudio não sofreu no entanto nenhum impacto durante a realização desta série de testes.

Resumindo as principais conclusões:

- existe uma relação directa entre QoS e QoE;
- a taxa de perda de pacotes revela-se o parâmetro com maior efeito na qualidade de experiência entre os parâmetros testados;
- uma taxa de perdas superior a 10% tem um grande efeito de degradação na qualidade de experiência;
- uma largura de banda inferior a 50kbit/s não permitirá uma boa qualidade de vídeo;
- um atraso regular na comunicação entre emissor e receptor não afecta por si só a qualidade de vídeo;
- o áudio é resiliente às condições pouco favoráveis de uma rede, neste caso, usando a aplicação Skype (poderão existir algoritmos de correcção próprios da aplicação);
- quanto menor for a quantidade de informação a ser transmitida maior será a resistência à degradação da QoE provocada pelas condições não favoráveis;
- o *jitter* possui grande influência na qualidade de experiência verificada;
- um valor de *jitter* de 74ms é o suficiente para produzir uma qualidade de vídeo reduzida;

- o áudio de uma videochamada consegue manter uma qualidade estável face ao jitter até um valor aproximadamente de 300ms;
- uma grande carga de tráfego TCP não produzirá efeitos tão acentuados na deterioração da qualidade de vídeo como o tráfego UDP;
- a qualidade áudio de uma videochamada possui uma boa imunidade a grandes cargas de tráfego, tanto de tráfego TCP como UDP.

6.1 Trabalho Futuro

O trabalho executado no âmbito desta tese limitou-se ao uso da aplicação Skype para a realização das videochamadas. Um trabalho deste tipo beneficiaria do uso de várias aplicações deste tipo (Windows Live Messenger, Facetime) para repetir a série de testes efectuados neste trabalho. Desta forma obteria-se mais um elemento de comparação que nos permitiria observar também a existência de diferenças entre as aplicações no que se refere à sua maneira de lidar com condições de rede adversas.

Como é mencionado no conteúdo desta dissertação existem também vários modelos de avaliação de vídeo e áudio, existindo também outras ferramentas não livres que executam este mesmo tipo de avaliação vídeo e áudio. Executando testes com as mesmas características aos realizados neste trabalho, poderia-se comparar o desempenho de diversos modelos e ferramentas aplicados ao serviço de videochamada, revelando-se desta forma a adequabilidade de cada um. Para um estudo ainda mais completo seria interessante submeter os mesmos dados a uma avaliação humana constituída por um largo grupo de indivíduos estatisticamente relevante. Crê-se também que o estudo aqui apresentado produz resultados importantes que poderão permitir aprofundar a investigação e criação de eventuais modelos mapeando a relação QoS/QoE. Tais estudos provariam ser úteis para um acréscimo ao estudo de avaliação de satisfação por parte de um utilizador e para o efeito que as condições de rede produzem no QoE.

Bibliografia

- [1] Solange Rito Lima and Paulo Carvalho. Enabling self-adaptive qoe/qos control. In *Proceedings of the 2011 IEEE 36th Conference on Local Computer Networks*, LCN '11, pages 239–242, Washington, DC, USA, 2011. IEEE Computer Society.
- [2] M. Fiedler, T. Hossfeld, and Phuoc Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *Network, IEEE*, 24(2):36 –41, march-april 2010.
- [3] Jodi Forlizzi and Shannon Ford. The building blocks of experience: an early framework for interaction designers. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, DIS '00, pages 419–423, New York, NY, USA, 2000. ACM.
- [4] ITU-T Recommendation P.862. *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, February 2001. Recommendations of the ITU, Telecommunication Standardization Sector.
- [5] Margaret H Pinson and Stephen Wolf. A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 50(3):312–322, September 2004.
- [6] Stephen Wolf. Variable frame delay (vfd) parameters for video quality measurements. *NTIA Technical Memorandum TM-11-475*, April 2011.
- [7] Margaret H Pinson and Stephen Wolf. Video quality model for variable frame delay (vqm_vfd). *NTIA Technical Memorandum TM-11-482*, September 2011.
- [8] Om Malik. Skype by the numbers: It's really big. http://gigaom.com/2010/04/20/skype-_q4-_2009-_number, April 2010.

- [9] Leena Rao. Skype revenue up 20 percent to 860m in 2010; paid users up 19 percent. <http://techcrunch.com/2011/03/07/skype-revenue-up-20-percent-to-860m-in-2010-paid-users-up-19-percent/>, March 2011.
- [10] S. Winkler. Video quality measurement standards; current status and trends. In *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on*, pages 1–5, dec. 2009.
- [11] ITU-T Recommendation J.247. *Objective perceptual multimedia video quality measurement in the presence of a full reference*, August 2008. Recommendations of the ITU, Telecommunication Standardization Sector.
- [12] Yu-Chun Chang, Chi-Jui Chang, Kuan-Ta Chen, and Chin-Laung Lei. Radar chart: Scanning for high qoe in qos dimensions. In *Communications Quality and Reliability (CQR), 2010 IEEE International Workshop Technical Committee on*, pages 1–6, june 2010.
- [13] V. Menkovski, G. Exarchakos, A. Liotta, and A.C. Sanchez. Estimations and remedies for quality of experience in multimedia streaming. In *Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services (CENTRIC), 2010 Third International Conference on*, pages 11–15, aug. 2010.
- [14] M. Venkataraman, M. Chatterjee, and S. Chattopadhyay. Evaluating quality of experience for streaming video in real time. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pages 1–6, 30 2009-dec. 4 2009.
- [15] J. de Francisco. About the difference between qos and qoe. <http://consultaglobal.wordpress.com/2009/03/17/about-the-difference-between-qos-and-qoe-1/>, March 2009.
- [16] International Organization for Standardization ISO 9241-210. *Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*, 2008.
- [17] Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold P.O.S. Vermeeren, and Joke Kort. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the 27th international conference on Human factors in computing systems, CHI '09*, pages 719–728, New York, NY, USA, 2009. ACM.

- [18] Amir Mehmood. Understanding cross-layer effects on quality of experience for video over next generation mobile networks. Talk at Internet MRA Reunion Conference II, June 2011.
- [19] Tobias Hobetafeld, Michael Seufert, Matthias Hirth, Thomas Zinner, Phuoc Tran-Gia, and Raimund Schatz. Quantification of youtube qoe via crowdsourcing. In *Proceedings of the 2011 IEEE International Symposium on Multimedia, ISM '11*, pages 494–499, Washington, DC, USA, 2011. IEEE Computer Society.
- [20] ITU-T Recommendation P.800. *Methods for subjective determination of transmission quality*, August 1996. Recommendations of the ITU, Telecommunication Standardization Sector.
- [21] Q. Huynh-Thu and M. Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics Letters*, 44(13):800–801, 19 2008.
- [22] Stephen T. Welstead. *Fractal and Wavelet Image Compression Techniques*, pages 155-156. SPIE Press, 1999.
- [23] N. Thomos, N.V. Boulgouris, and M.G. Strintzis. Optimized transmission of jpeg2000 streams over wireless channels. *Image Processing, IEEE Transactions on*, 15(1):54–67, jan. 2006.
- [24] Video Quality Experts Group (VQEG). Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii. Available: <http://www.vqeg.org>, March 2003.
- [25] ITU-T Recommendation J.143. *User requirements for objective perceptual video quality measurements in digital cable television*, May 2000. Recommendations of the ITU, Telecommunication Standardization Sector.
- [26] ITU-T Recommendation BT.500. *Methodology for the subjective assessment of the quality of television pictures*, January 2012. Recommendations of the ITU, Telecommunication Standardization Sector.
- [27] ITU-T Recommendation P.910. *Subjective video quality assessment methods for multimedia applications*, April 2008. Recommendations of the ITU, Telecommunication Standardization Sector.
- [28] Margaret H. Pinson and Stephen Wolf. *Batch Video Quality Metric BVQM Users Manual*, September 2011. NTIA Handbook HB-11-441d.

- [29] Margaret H Pinson and Stephen Wolf. Reduced reference video calibration algorithms. *NTIA Report TR-06-433a*, July 2006.
- [30] Margaret H Pinson and Stephen Wolf. Reduced reference video calibration algorithms. *NTIA Report TR-07-433b*, November 2007.
- [31] ITU-T Recommendation J.244. *Full reference and reduced reference calibration methods for video transmission systems with constant misalignment of spatial and temporal domains with constant gain and offset*, April 2008. Recommendations of the ITU, Telecommunication Standardization Sector.
- [32] Margaret H Pinson and Stephen Wolf. Low bandwidth reduced reference video quality monitoring system. *Proc. First International Workshop on Video Processing and Quality Metrics for Consumer electronics*, January 2005.
- [33] Margaret H Pinson and Stephen Wolf. Video quality measurement techniques. *NTIA Report 02-392*, June 2002.
- [34] Stephen Wolf. A full reference (fr) method using causality processing for estimating variable video delays. *NTIA Technical Memorandum TM-10-463*, October 2009.
- [35] Marta Carbone and Luigi Rizzo. Dummynet revisited. *ACM SIGCOMM Computer Communication Review*, 40(2) pg.12-20, March 2010.
- [36] A. Botta, A. Dainotti, and A. Pescapè. A tool for the generation of realistic network workload for emerging networking scenarios. *Computer Networks (Elsevier)*, 2012.
- [37] Splitcam. <http://www.splitcamera.com>. Consulted in December 2011.
- [38] Supertintin. <http://www.supertintin.com/>. Consulted in December 2011.
- [39] Virtualdub. <http://www.virtualdub.org/>. Consulted in March 2012.
- [40] The consumer digital video library. http://www.cdvl.org/find_videos/details.php?id=81. Consulted in February 2012.
- [41] Yuv video sequences. <http://trace.eas.asu.edu/yuv/>. Consulted in February 2012.
- [42] The open speech repository. http://www.voiptroubleshooter.com/open_speech/american/OSR_us_000_0031_8k.wav. Consulted in March 2012.

- [43] NTIA/ITS. Batch video quality metric (bvqm) software. <http://vqm.its.bldrdoc.gov/>. Consulted in January 2012.
- [44] NLANR/DAST. Iperf. <http://iperf.fr/>. Consulted in April 2012.

BIBLIOGRAFIA

Anexo A

Resultados individuais da avaliação de vídeo

A.1 Loss Ratio vs Delay

LR%	Delay ms											
	0			50			100			150		
0	0,1442	0,1452	0,1393	0,1431	0,1399	0,137	0,1446	0,1465	0,1429	0,1466	0,1499	0,1451
	0,1429			0,14			0,144666667			0,1472		
5	0,2928	0,2413	0,629	0,2079	0,1864	0,4954	0,2602	0,313	0,4109	0,2531	0,2586	0,348
	0,3877			0,296566667			0,328033333			0,286566667		
10	0,4606	0,6377	0,5435	0,5322	0,5877	0,5579	0,354	0,3568	0,4131	0,5046	0,3168	0,2555
	0,547266667			0,559266667			0,374633333			0,358966667		
15	0,7108	0,7592	0,7825	0,7249	0,8099	0,6976	0,7575	0,781	0,7733	0,7197	0,7724	0,7736
	0,750833333			0,744133333			0,7706			0,755233333		
20	0,7697	0,7853	0,785	0,7006	0,6503	0,7767	0,7853	0,7936	0,679	0,781	0,781	0,8043
	0,78			0,7092			0,752633333			0,788766667		

Figura A.1: Vídeos "chart" - Pontuações VQM_VFD - LR vs Delay

LR	Delay				
	0	50	100	150	200
0	0,1206	0,12	0,1203	0,128	0,1202
5	0,2099	0,4322	0,2425	0,1388	0,2752
10	0,432	0,3634	0,1793	0,5279	0,4999
15	0,6607	0,6858	0,7486	0,7904	0,735
20	0,7344	0,7034	0,8016	0,6955	0,795

Figura A.2: Vídeos "train" - Pontuações VQM_VFD - LR vs Delay

A.2 Loss Ratio vs Bandwidth

LR%	Bw Kbit/s														
	unl			100			75			50			25		
0	0,1442	0,1452	0,1393	0,3967	0,3322	0,3463	0,5516	0,3945	0,4145	0,5045	0,6908	0,6868	0,7691	0,7416	0,7484
	0,1429			0,3584			0,453533333			0,627366667			0,753033333		
5	0,5321	0,4501	0,3921	0,4707	0,4846	0,4726	0,5738	0,5648	0,5092	0,6638	0,6099	0,5892	0,7507	0,7586	0,7958
	0,4581			0,475966667			0,549266667			0,620966667			0,768366667		
10	0,5223	0,4571	0,4582	0,5723	0,5304	0,5419	0,5348	0,6231	0,6069	0,7156	0,6482	0,6881	0,7905	0,7385	0,7678
	0,4792			0,5482			0,588266667			0,683966667			0,7656		
15	0,6595	0,7696	0,7407	0,7941	0,796	0,7172	0,7114	0,7619	0,7392	0,6733	0,7634	0,775	0,7491	0,814	0,7346
	0,723266667			0,7691			0,7375			0,737233333			0,7659		
20	0,5671	0,8025	0,7014	0,8034	0,8581	0,7397	0,7521	0,7491	0,76	0,8797	0,8797	0,7822	0,766	0,8355	0,7665
	0,690333333			0,8004			0,753733333			0,8472			0,789333333		

Figura A.3: Vídeos "chart" - Pontuações VQM_VFD - LR vs BW

LR	Bandwidth					
	unl	100	75	50	25	
0	0,1206	0,41	0,5385	0,6391	0,7393	
5	0,1914	0,6459	0,6918	0,7148	0,7495	
10	0,6132	0,611	0,7072	0,7168	0,7422	
15	0,7545	0,6984	0,7769	0,6972	0,7471	
20	0,793	0,7582	0,7615	0,7214	0,8122	

Figura A.4: Vídeos "train" - Pontuações VQM_VFD - LR vs BW

A.3 Bandwidth vs Delay

Delay ms	Bw Kbit/s														
	unl			100			75			50			25		
0	0,1442	0,1452	0,1393	0,4024	0,3847	0,41	0,4864	0,4409	0,4448	0,5279	0,5297	0,5574	0,7207	0,7163	0,7331
	0,1429			0,399033333			0,457366667			0,538333333			0,723366667		
100	0,1425	0,1443	0,1465	0,4109	0,4002	0,3779	0,6045	0,4378	0,4306	0,6285	0,674	0,5195	0,7649	0,7181	0,753
	0,144433333			0,396333333			0,490966667			0,607333333			0,745333333		
200	0,149	0,1409	0,1456	0,4201	0,3756	0,3959	0,4062	0,4164	0,4622	0,5507	0,5542	0,5334	0,777	0,7393	0,7136
	0,145166667			0,3972			0,428266667			0,5461			0,7433		
300	0,1364	0,1348	0,1344	0,4596	0,3868	0,3622	0,4191	0,4712	0,4025	0,5182	0,5301	0,5286	0,6941	0,6656	0,6605
	0,1352			0,402866667			0,430933333			0,525633333			0,6734		
400	0,1409	0,1444	0,143	0,4087	0,3679	0,3517	0,4188	0,4223	0,3999	0,5269	0,5269	0,5082	0,6951	0,7171	0,7208
	0,142766667			0,3761			0,413666667			0,520666667			0,711		

Figura A.5: Vídeos "chart" - Pontuações VQM_VFD - BW vs Delay

Delay ▾	Bandwidth				
	unl ▾	100 ▾	75 ▾	50 ▾	25 ▾
0	0,1206	0,55	0,5798	0,6793	0,7301
100	0,2336	0,4594	0,5515	0,6375	0,7257
200	0,6961	0,5599	0,5939	0,6634	0,7294
300	0,5192	0,6473	0,6166	0,6478	0,7227
400	0,5839	0,6931	0,6809	0,6995	0,7216

Figura A.6: Vídeos "train" - Pontuações VQM_VFD - BW vs Delay

A.4 Jitter

Jitter 50 100 150 200		Jitter 100 200 300 400		Jitter 100 300 700 900	
	vqm_vfd		vqm_vfd		vqm_vfd
chart_jitter_proc1	0,6018	chart_jitter_proc1	0,7125	chart_jitter_proc1	0,7523
chart_jitter_proc2	0,6342	chart_jitter_proc2	0,7692	chart_jitter_proc2	0,6977
chart_jitter_proc3	0,5807	chart_jitter_proc3	0,7731	chart_jitter_proc3	0,72
Medições de jitter (ms)		Medições de jitter (ms)		Medições de jitter (ms)	
jitter	59.353	jitter	104.683	jitter	334.139
datagrams out of order	565	datagrams out of order	599	datagrams out of order	582
jitter	77.432	jitter	112.418	jitter	235.559
datagrams out of order	548	datagrams out of order	595	datagrams out of order	598
jitter	81.122	jitter	123.927	jitter	428.121
datagrams out of order	535	datagrams out of order	588	datagrams out of order	571
jitter	86.897	jitter	127.596	jitter	182.453
datagrams out of order	552	datagrams out of order	616	datagrams out of order	573
jitter	68.668	jitter	171.047	jitter	324.436
datagrams out of order	559	datagrams out of order	585	datagrams out of order	601

Figura A.7: Vídeos "chart" - Pontuações VQM_VFD e medições de jitter

Jitter 50 100 150 200		Jitter 100 200 300 400		Jitter 100 300 700 900	
	vqm_vfd		vqm_vfd		vqm_vfd
train_jitter_proc1	0,6471	train_jitter_proc1	0,6484	train_jitter_proc1	0,6778
train_jitter_proc2	0,6596	train_jitter_proc2	0,6811	train_jitter_proc2	0,6835
train_jitter_proc3	0,6181	train_jitter_proc3	0,695	train_jitter_proc3	0,7315
Medições de jitter (ms)		Medições de jitter (ms)		Medições de jitter (ms)	
jitter	59.353	jitter	104.683	jitter	334.139
datagrams out of order	565	datagrams out of order	599	datagrams out of order	582
jitter	77.432	jitter	112.418	jitter	235.559
datagrams out of order	548	datagrams out of order	595	datagrams out of order	598
jitter	81.122	jitter	123.927	jitter	428.121
datagrams out of order	535	datagrams out of order	588	datagrams out of order	571
jitter	86.897	jitter	127.596	jitter	182.453
datagrams out of order	552	datagrams out of order	616	datagrams out of order	573
jitter	68.668	jitter	171.047	jitter	324.436
datagrams out of order	559	datagrams out of order	585	datagrams out of order	601

Figura A.8: Vídeos "train" - Pontuações VQM_VFD e medições de jitter

A.5 TCP

Vídeo_PacotesTCP por seg.	vqm_vfd
chart_notraffic_1	0,1484
chart_notraffic_2	0,1409
chart_notraffic_3	0,1421
chart_80_1	0,3313
chart_80_2	0,3655
chart_80_3	0,3097
chart_300_1	0,4
chart_300_2	0,2875
chart_300_3	0,1679
chart_500_1	0,5285
chart_500_2	0,4417
chart_500_3	0,1776
chart_1000_1	0,2853
chart_1000_2	0,2568
chart_1000_3	0,4089

Figura A.9: Vídeos "chart" - Pontuações VQM_VFD - TCP

Vídeo_PacotesTCP por seg.	vqm_vfd
train_notraffic_1	0,1154
train_notraffic_2	0,1118
train_notraffic_3	0,127
train_80_1	0,2251
train_80_2	0,2936
train_80_3	0,139
train_300_1	0,5366
train_300_2	0,4665
train_300_3	0,3065
train_500_1	0,2748
train_500_3	0,5193
train_500_5	0,4824
train_1000_1	0,4362
train_1000_2	0,1998
train_1000_3	0,2902

Figura A.10: Vídeos "train" - Pontuações VQM_VFD - TCP

A.6 UDP

Video_PacotesUDP por seg.	vqm_vfd
chart_notraffic_1	0,1484
chart_notraffic_2	0,1409
chart_notraffic_3	0,1421
chart_20_1	0,3067
chart_20_2	0,3509
chart_20_3	0,295
chart_50_1	0,5712
chart_50_2	0,374
chart_50_3	0,2971
chart_80_1	0,4253
chart_80_2	0,5751
chart_80_3	0,3241
chart_300_1	0,551
chart_300_2	0,6565
chart_300_3	0,5538
chart_1000_1	0,4718
chart_1000_2	0,5476
chart_1000_3	0,5427

Figura A.11: Vídeos "chart" - Pontuações VQM_VFD - UDP

Video_PacotesUDP por seg.	vqm_vfd
train_notraffic_1	0,1154
train_notraffic_2	0,1118
train_notraffic_3	0,127
train_20_3	0,1584
train_20_4	0,2531
train_20_5	0,4762
train_50_1	0,2857
train_50_2	0,3041
train_50_3	0,5875
train_80_3	0,5925
train_80_4	0,6636
train_80_5	0,6231
train_300_3	0,7318
train_300_4	0,7365
train_300_5	0,6645
train_1000_1	0,5892
train_1000_2	0,7376
train_1000_3	0,6381

Figura A.12: Vídeos "train" - Pontuações VQM_VFD - UDP

ANEXO A. RESULTADOS INDIVIDUAIS DA AVALIAÇÃO DE VÍDEO

Anexo B

Resultados individuais da avaliação de áudio

B.1 Loss Ratio vs Delay

LR%	Delay ms														
	0			50			100			150			200		
0	2,163	2,062	2,18	2,018	2,265	2,041	1,926	2,029	1,772	2,014	1,975	2,058	1,839	1,828	1,793
	2,135			2,108			1,909			2,015666667			1,82		
5	1,999	2,048	2,03	2,059	1,95	1,937	1,481	1,338	1,562	2,088	1,866	1,887	1,756	1,852	1,689
	2,025666667			1,982			1,460333333			1,947			1,765666667		
10	1,525	1,742	1,874	1,854	1,628	1,858	1,799	1,934	1,747	1,889	1,77	1,893	1,93	1,903	1,827
	1,713666667			1,78			1,826666667			1,850666667			1,886666667		
15	1,266	1,343	1,203	1,312	1,277	1,387	1,322	1,368	1,248	1,186	1,29	1,304	1,235	1,292	1,216
	1,270666667			1,325333333			1,312666667			1,26			1,247666667		
20	1,237	1,2	1,268	1,535	1,261	1,193	1,289	1,217	1,307	1,209	1,204	1,206	1,301	1,318	1,293
	1,235			1,329666667			1,271			1,206333333			1,304		

Figura B.1: Faixas áudio "chart" - Pontuações MOS - LR vs Delay

LR	Delay				
	0	50	100	150	200
0	2,566	2,515	2,559	2,169	1,723
5	2,114	1,546	2,209	2,153	2,317
10	2,142	2,291	2,101	2,198	2,074
15	1,447	1,664	1,575	1,509	1,608
20	1,488	1,818	1,655	1,663	1,668

Figura B.2: Faixas áudio "train" - Pontuações MOS - LR vs Delay

B.2 Loss Ratio vs Bandwidth

LR%	Bw Kbit/s														
	unl			100			75			50			25		
0	2,163	2,062	2,18	1,671	1,531	1,686	1,656	1,994	1,36	1,513	1,337	1,716	2,456	1,611	1,798
	2,135			1,629333333			1,67			1,522			1,955		
5	1,405	1,703	1,987	1,58	1,573	1,403	1,448	1,646	1,588	1,854	1,401	1,341	1,351	1,549	1,353
	1,698333333			1,518666667			1,560666667			1,532			1,417666667		
10	1,877	1,553	1,742	1,622	1,255	1,577	1,518	1,459	1,49	2,023	1,52	1,359	1,271	1,646	1,394
	1,724			1,484666667			1,489			1,634			1,437		
15	1,226	1,274	1,53	1,533	1,504	1,195	1,307	1,667	1,649	1,428	1,637	1,331	2,368	1,234	1,609
	1,343333333			1,410666667			1,541			1,465333333			1,737		
20	2,255	1,193	1,524	1,441	1,921	1,764	1,17	1,425	1,788	1,348	1,268	1,732	1,346	2,026	1,353
	1,657333333			1,708666667			1,461			1,449333333			1,575		

Figura B.3: Faixas áudio "chart" - Pontuações MOS - LR vs BW

LR	Bandwidth				
	unl	100	75	50	25
0	2,566	2,283	2,587	2,703	2,145
5	2,407	2,497	1,986	1,863	1,52
10	2,63	2,244	1,775	1,526	1,959
15	1,682	1,807	1,452	1,746	1,529
20	1,703	1,873	1,629	1,443	2,151

Figura B.4: Faixas áudio "train" - Pontuações MOS - LR vs BW

B.3 Bandwidth vs Delay

Delay ms	Bw Kbit/s														
	unl			100			75			50			25		
0	2,163	2,062	2,18	2,34	2,145	2,101	1,627	1,659	1,737	1,478	1,673	1,674	1,332	1,474	1,662
	2,135			2,195333333			1,674333333			1,608333333			1,489333333		
100	2,009	2,157	2,309	1,813	1,614	1,997	1,639	1,598	1,607	2,059	1,827	1,495	2,197	1,797	2,096
	2,158333333			1,808			1,614666667			1,793666667			2,03		
200	2,156	2,229	2,18	1,661	1,807	1,905	1,487	1,871	1,858	1,772	2,103	1,584	2,424	2,28	1,49
	2,188333333			1,791			1,738666667			1,819666667			2,064666667		
300	2,134	2,165	2,174	2,346	1,649	1,972	1,856	2,429	1,733	1,63	1,5	1,397	1,795	1,717	1,335
	2,157666667			1,989			2,006			1,509			1,615666667		
400	2,179	2,201	1,946	1,854	1,913	2,207	1,828	2,051	1,858	1,346	1,494	1,56	1,376	1,48	1,896
	2,108666667			1,991333333			1,912333333			1,466666667			1,584		

Figura B.5: Faixas áudio "chart" - Pontuações MOS - BW vs Delay

Delay ▼	Bandwidth				
	unl ▼	100 ▼	75 ▼	50 ▼	25 ▼
0	2,566	2,25	2,696	2,595	2,72
100	2,829	2,579	2,353	2,38	1,452
200	2,895	1,691	2,798	2,453	2,239
300	2,391	2	2,778	2,838	1,578
400	3,093	2,632	2,438	2,747	2,897

Figura B.6: Faixas áudio "train" - Pontuações MOS - BW vs Delay

B.4 Jitter

Jitter 50 100 150 200	MOS		Jitter 100 200 300 400	MOS		Jitter 100 300 700 900	MOS
chart_jitter_proc1	2.509		chart_jitter_proc1	2.132		chart_jitter_proc1	1.203
chart_jitter_proc2	2.296		chart_jitter_proc2	2.004		chart_jitter_proc2	1.354
chart_jitter_proc3	2.099		chart_jitter_proc3	1.668		chart_jitter_proc3	1.665

Figura B.7: Faixas áudio "chart" - Pontuações MOS - Jitter

Jitter 50 100 150 200	MOS		Jitter 100 200 300 400	MOS		Jitter 100 300 700 900	MOS
train_jitter_proc1	2.934		train_jitter_proc1	3.089		train_jitter_proc1	1.887
train_jitter_proc2	2.529		train_jitter_proc2	2.752		train_jitter_proc2	1.341
train_jitter_proc3	2.551		train_jitter_proc3	2.218		train_jitter_proc3	1.215

Figura B.8: Faixas áudio "train" - Pontuações MOS - Jitter

B.5 TCP

Áudio_PacotesTCP por seg.	MOS
chart_notraffic_1	2,26
chart_notraffic_2	2,255
chart_notraffic_3	2,288
chart_80_1	2,293
chart_80_2	2,48
chart_80_3	2,45
chart_300_1	2,118
chart_300_2	2,048
chart_300_3	2,522
chart_500_1	2,087
chart_500_2	2,183
chart_500_3	2,067
chart_1000_1	2,175
chart_1000_2	2,124
chart_1000_3	2,31

Figura B.9: Faixas áudio "chart" - Pontuações MOS - TCP

Áudio_PacotesTCP por seg.	MOS
train_notraffic_1	2,432
train_notraffic_2	2,324
train_notraffic_3	2,624
train_80_1	2,563
train_80_2	2,335
train_80_3	2,294
train_300_1	2,808
train_300_2	2,509
train_300_3	2,459
train_500_1	2,392
train_500_2	2,561
train_500_3	2,25
train_1000_1	2,34
train_1000_2	2,5
train_1000_3	2,332

Figura B.10: Faixas áudio "train" - Pontuações MOS - TCP

B.6 UDP

Áudio_PacotesUDP por seg.	MOS
chart_notraffic_1	2,26
chart_notraffic_2	2,255
chart_notraffic_3	2,288
chart_20_1	2,135
chart_20_2	2,154
chart_20_3	2,339
chart_50_1	2,214
chart_50_2	1,95
chart_50_3	2,178
chart_80_1	2,393
chart_80_2	2,213
chart_80_3	2,128
chart_300_1	2,324
chart_300_2	2,445
chart_300_3	2,331
chart_1000_1	2,209
chart_1000_2	1,953
chart_1000_3	2,368

Figura B.11: Faixas áudio "chart" - Pontuações MOS - UDP

Áudio_PacotesUDP por seg.	MOS
train_notraffic_1	2,432
train_notraffic_2	2,324
train_notraffic_3	2,624
train_20_1	2,46
train_20_2	2,358
train_20_3	2,372
train_50_1	2,272
train_50_2	2,254
train_50_3	2,224
train_80_1	2,422
train_80_2	2,568
train_80_3	2,694
train_300_1	1,557
train_300_2	2,374
train_300_3	2,388
train_1000_1	2,816
train_1000_2	2,595
train_1000_3	2,721

Figura B.12: Faixas áudio "train" - Pontuações MOS - UDP